2020

Structure-Function Analysis Of Bet Proteins In Transcription

Michael Thomas Werner
University of Pennsylvania

Structure-Function Analysis Of Bet Proteins In Transcription

Abstract

The widely expressed bromodomain and extraterminal motif (BET) proteins bromodomain-containing protein 2 (BRD2), BRD3, and BRD4 are multifunctional transcriptional regulators that bind acetylated chromatin via their conserved tandem bromodomains. Small molecules that target BET bromodomains are being tested for various diseases, but typically do not discern between BET family members. Genomic distributions and protein partners of BET proteins have been described, but the basis for differences in BET protein function within a given lineage remains unclear. By establishing a gene knockout–rescue system in a Brd2-null erythroblast cell line, we compared a series of mutant and chimeric BET proteins for their ability to modulate cell growth, differentiation and gene expression. We found that the BET N-terminal halves bearing the bromodomains convey marked differences in protein stability, but do not account for specificity in BET protein function. Instead, when BET proteins were expressed at comparable levels, their specificity was largely determined by the C-terminal half. Remarkably, a chimeric BET protein comprised of the N-terminal half of the structurally similar short BRD4 isoform (BRD4S) and the C-terminal half of BRD2 functioned similarly to intact BRD2. We traced part of the BRD2-specific activity to a previously uncharacterized short segment predicted to harbor a coiled coil (CC) domain. Deleting the CC segment impaired BRD2's ability to restore growth and differentiation, and the CC region functioned in conjunction with the adjacent ET domain (ETCC) to impart BRD2-like activity onto BRD4S. We found that despite having different functions, BET proteins share similar genome-wide chromatin binding profiles. Neither exchanging the bromodomain-containing N-terminal nor the ETCC-bearing C-terminal regions of BET proteins alters their distribution. Rather, BRD2-specific function may in part be mediated by ETCC interactions with the RNA polymerase II associated factor and casein kinase II complexes. In summary, our results identify distinct BET protein domains that regulate protein turnover and biological activities.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Immunology

First Advisor

Gerd A. Blobel

Keywords

BET proteins, Chromatin biology, Erythropoiesis, Gene regulation, Transcription

Subject Categories

Biology | Genetics | Molecular Biology

STRUCTURE-FUNCTION ANALYSIS OF BET PROTEINS IN TRANSCRIPTION

Michael T. Werner

A DISSERTATION

in

Immunology

Presented to the Faculties of the University of Pennsylvania

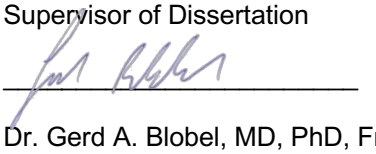
in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation



Dr. Gerd A. Blobel, MD, PhD, Frank E. Weise III Professor of Pediatrics

Graduate Group Chairperson

Dr. David M. Allman, PhD, Professor of Pathology and Laboratory Medicine

Dissertation Committee

Dr. Warren S. Pear, MD, PhD, Gaylord P. and Mary Louise Harnwell Professor

Dr. Shelley L. Berger, PhD, Daniel S. Och University Professor

Dr. Golnaz Vahedi, PhD, Assistant Professor of Genetics

Dr. Irfan A. Asangani, PhD, Assistant Professor of Cancer Biology

ACKNOWLEDGMENTS

I would like to thank many people for their assistance, mentorship, and support over the last five years. First, I will thank my parents, William and Linda Werner, for encouraging my ambitions early in my childhood and for remaining supportive throughout. I also owe a large debt of gratitude to my thesis mentor, Dr. Gerd Blobel. It has been a pleasure discussing science with him over the last several years, and I am grateful that he welcomed me into his lab and imparted his wisdom and direction. I would also like to thank him for recruiting a diverse and talented group of young scientists into the lab that I have had the pleasure to work with and learn from. The following Blobel lab members all made significant contributions to my research: Hongxin Wang, Sarah Hsu, Nicole Hamigami, Jennifer Yano, Aaron Stonestrom, Vivek Behera, Marit Vermut, Carolyne Face, Laavanya Sankaranarayanan, Perry Evans, and Zhe (Jim) Zhang. I also benefited from collaborations with Cheryl Keller, Belinda Giardine, and Ross Hardison of the Hardison Lab and with Joel Mackay. The following members of my thesis committee have also been a tremendous resource for scientific and career advice: Drs. Golnaz Vahedi, Irfan Asangani, Shelley Berger, and Warren Pear. Completion of my graduate school education would not have been possible without the administrative and moral support provided by Dr. Skip Brass and Maggie Krall of the Medical Scientist Training Program and Mary Taylor and Dr. David Allman of the Immunology Graduate Group. Finally, I'll conclude by thanking my fiancé, Melanie Kier, for her unconditional support.

ABSTRACT

STRUCTURE-FUNCTION ANALYSIS OF BET PROTEINS IN TRANSCRIPTION

Michael T. Werner

Gerd A. Blobel

The widely expressed bromodomain and extraterminal motif (BET) proteins bromodomain-containing protein 2 (BRD2), BRD3, and BRD4 are multifunctional transcriptional regulators that bind acetylated chromatin via their conserved tandem bromodomains. Small molecules that target BET bromodomains are being tested for various diseases, but typically do not discern between BET family members. Genomic distributions and protein partners of BET proteins have been described, but the basis for differences in BET protein function within a given lineage remains unclear. By establishing a gene knockout–rescue system in a *Brd2*-null erythroblast cell line, we compared a series of mutant and chimeric BET proteins for their ability to modulate cell growth, differentiation and gene expression. We found that the BET N-terminal halves bearing the bromodomains convey marked differences in protein stability, but do not account for specificity in BET protein function. Instead, when BET proteins were expressed at comparable levels, their specificity was largely determined by the C-terminal half. Remarkably, a chimeric BET protein comprised of the N-terminal half of the structurally similar short BRD4 isoform (BRD4S) and the C-terminal half of BRD2 functioned similarly to intact BRD2. We traced part of the BRD2-specific activity to a previously uncharacterized short segment predicted to harbor a coiled coil (CC) domain. Deleting the CC segment impaired BRD2's ability to restore growth and differentiation, and the CC region functioned in conjunction with the adjacent ET domain (ETCC) to impart BRD2-like

activity onto BRD4S. We found that despite having different functions, BET proteins share similar genome-wide chromatin binding profiles. Neither exchanging the bromodomain-containing N-terminal nor the ETCC-bearing C-terminal regions of BET proteins alters their distribution. Rather, BRD2-specific function may in part be mediated by ETCC interactions with the RNA polymerase II associated factor and casein kinase II complexes. In summary, our results identify distinct BET protein domains that regulate protein turnover and biological activities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

PREFACE

A large fraction of the work I present in this dissertation is published (Werner et al)[1]. The remainder of this work will form the basis of future studies.

The first chapter of this dissertation is a broad introduction to serve as backdrop for the biological questions that I address about BET proteins. The second chapter includes detailed materials and methods used in these studies. The third and fourth chapters serve as result sections for my two major lines of investigation: 1) isolating the critical element that distinguishes BET proteins from each other and 2) characterizing this domain. These chapters contain focused introductions, experimental results, and tailored discussions. The final chapter draws on these findings to make broader conclusions and to discuss the implications, limitations, and future directions of this work.

CHAPTER 1: INTRODUCTION

Gene transcription is regulated to ensure that cell-appropriate genes are expressed at adequate levels. For example, red blood cells must express globin genes in order to produce oxygen-carrying hemoglobin. They do this while simultaneously preventing the expression of unnecessary genes. Such cell-type specific gene expression allows each cell and tissue to take on specialized functions – whether carrying oxygen throughout the body or conveying electrical impulses across long distances in the nervous system. A breakdown in transcriptional regulation can result in disease by interfering with these biological processes. Elaborate regulatory mechanisms to control gene expression have thus evolved. In ancestral organisms such as bacteria, these mechanisms can be thought of as relatively simple transcriptional circuits. A classic example is the lac operon, which turns on lactose metabolism genes only when bacteria are in the presence of lactose but not glucose. In multicellular eukaryotes, the regulatory mechanisms are more complicated but achieve the same goal: finely tuned expression of genes in the right cell and at the right time.

The goal of the work I present here is to describe one layer of eukaryotic transcriptional regulation. This layer falls within the realm of epigenetics, a broad field of study that describes changes in phenotypes that are unrelated to changes in the basic genetic code. Though the term was originally used to describe inheritance patterns that could not be explained by typical genetic mechanisms,

the term is now applied more broadly to the study of transcriptional regulation and chromatin biology.

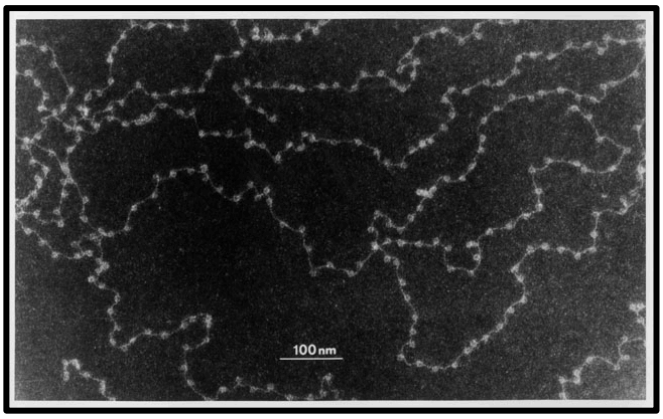


Figure 1.1. Electron micrograph of chromatin. Imaged was produced by Olins and Olins who first published the finding in 1974[2]. The image depicts the "beads on a string" model of chromatin structure. DNA wraps around regularly spaced histones. Each DNA-histone core is a nucleosome, the fundamental unit of chromatin and a major regulator of gene expression.

Chromatin is the dense meshwork of DNA and protein found within the nucleus. In its most compact form, chromatin is organized into ~30μm width structures called chromosomes. Packaging of DNA into chromosomes is critical for the physical delivery of the genome to daughter cells during mitosis and meiosis. When cells are not dividing (interphase), chromatin is more loosely organized into smaller structures. High resolution imaging of interphase chromatin reveals a "beads on a string" pattern (Figure 1.1). This corresponds to the periodic wrapping of DNA (the string) around core histone proteins (the beads). A significant amount of biochemical and functional characterization has revealed that the structure and chemical modification of chromatin at this level is critical to the regulation of transcription[3]. Thus, whereas the term *genetics* can be applied to the

study of elements and diseases hard coded into the sequence of DNA itself and inherited through cell division, the term *epigenetics* can be thought of as the study of how the structure of chromatin controls the expression of genes and how dysregulation manifests as disease.

In the remainder of this chapter, I will introduce a family of proteins, bromodomain and extraterminal motif (BET) proteins, that are broadly considered to be “epigenetic reader” proteins. This terminology is based on the finding that these proteins possess specialized domains that recognize signatures of chemically modified chromatin. BET proteins thus “read” the epigenome by translating these chemical signals into transcriptional output.

BET proteins: linking chromatin acetylation to transcription

Common post-translational modifications of proteins include phosphorylation, acetylation and methylation, which are all well known to affect protein function. Histones, in fact, were the first proteins identified to be altered by acetylation and methylation, and even in these initial studies the functional consequence of acetylation and methylation on RNA synthesis was appreciated[4]. This early work provided the first evidence that post-translational modifications of chromatin may influence transcription.

Chromatin acetylation is associated with active transcription

Early characterization of histone acetylation revealed that the small molecule sodium butyrate could increase histone acetylation and induce cancer cell differentiation, providing an additional link between histone acetylation and gene expression[5]. Subsequent studies realized sodium butyrate targeted histone

deacetylase activity found in nuclear protein fractions and that in addition to increasing histone acetylation levels it also altered chromatin accessibility[6,7]. A direct link between the acetylation of histone tails at lysine residues and transcription was not tested until several decades later when antibodies raised against acetylated histone isoforms were used to immunoprecipitated chromatin containing active but not inactive genes[8]. Immunofluorescent staining revealed non-uniform distribution of acetylated histone isoforms into distinct domains associated with different levels of transcriptional activity on polytene chromosomes of Drosophila salivary glands[9]. Together, these studies linked histone acetylation and chromatin decondensation (based on imaging of the polytene chromosomes) with transcriptional activity. Purification of the first histone acetyltransferases (KAT2A/GCN5) and deacetylases (HDAC1/RPD3), known transcriptional activators and repressors in yeast respectively, further cemented the notion that the directed modification of chromatin by acetylation was directly linked to transcription[10,11].

The bromodomain is a specialized domain that binds to acetylated residues

How does chromatin acetylation activate transcription? It was first proposed that acetylation of lysines may neutralize the net positive charge of histones[12]. It was thought that the negative charge of the acetyl group would relax the interaction between positively charged histones and negatively charged DNA, leading to local decondensation of chromatin and access of the underlying DNA to various transcription factors. However, in addition to any biochemical affect acetylation may have on the structure of chromatin, a common ~110 amino acid sequence

possessed by many chromatin binding proteins was identified as an acetyl-lysine recognition domain. This domain was termed the bromodomain. A solution structure of the bromodomain from the acetyltransferase PCAF/KAT2B revealed the molecular interface of the acetyl-lysine recognition cleft of this domain with acetylated lysine[13]. Despite a considerable amount of diversity in the sequences of various bromodomains, this acetyl-lysine recognition pocket is strikingly conserved across each bromodomains[14]. Bromodomains are found in a multitude of chromatin-associated proteins, including histone acetyltransferases, methyltransferases, chromatin remodeling proteins, nuclear scaffolds and helicases[15]. These transcriptional regulators can be classified as epigenetic writers (histone acetyltransferases), erasers (histone deacetylases), or readers (histone acetylation binding proteins) based on their ability to add, remove, or bind to these chemical modifications on chromatin (Figure 1.2). Thus, the prevailing hypothesis as to how acetylation of chromatin affects transcription is that in addition to any electrostatic changes that acetyl groups may convey, acetyl groups also recruit a network of acetylation-dependent transcriptional regulatory proteins onto chromatin[16].

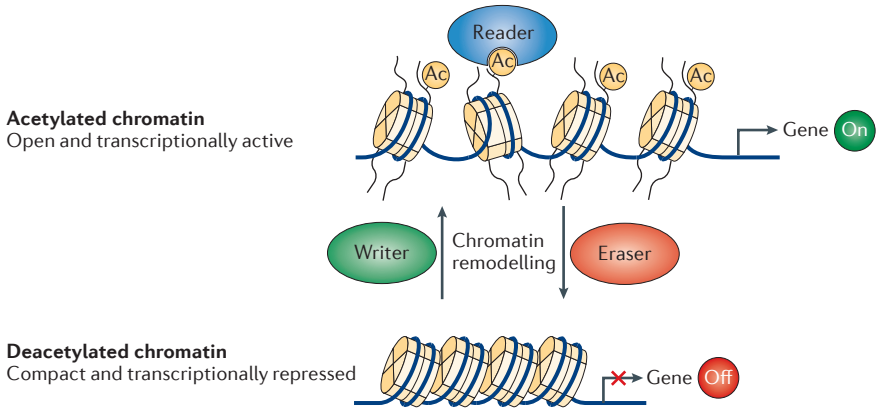


Figure 1.2. Schematic depicting the correlation between chromatin acetylation and active transcription. Imaged is adapted from Verdin and Ott (2015)[17]. Active gene expression is associated with acetylated nucleosomes. Enzymes that acetylate chromatin (histone acetylases) are termed writers, whereas those that remove the mark (histone deacetylases) are termed erasers. Chromatin remodeling proteins increase the spacing between adjacent histones. Proteins that recognize the acetylated but not unacetylated nucleosomes are termed readers.

Discovery of BET proteins and their association with transcription

The human genome contains 56 individual bromodomains contained within 42 distinct proteins[15]. Sequence homology of the bromodomains can be used to study the evolutionary relationships between these bromodomain-containing proteins (Figure 1.3). This analysis reveals a lineage of proteins – BRD2, BRD3, BRD4 and BRDT – each possessing two bromodomains in the N-terminal halves of the proteins in addition to a conserved motif termed the extraterminal domain (ET domain) situated C-terminal to the second bromodomain. These proteins are thus named bromodomain and extraterminal motif (BET) proteins[18].

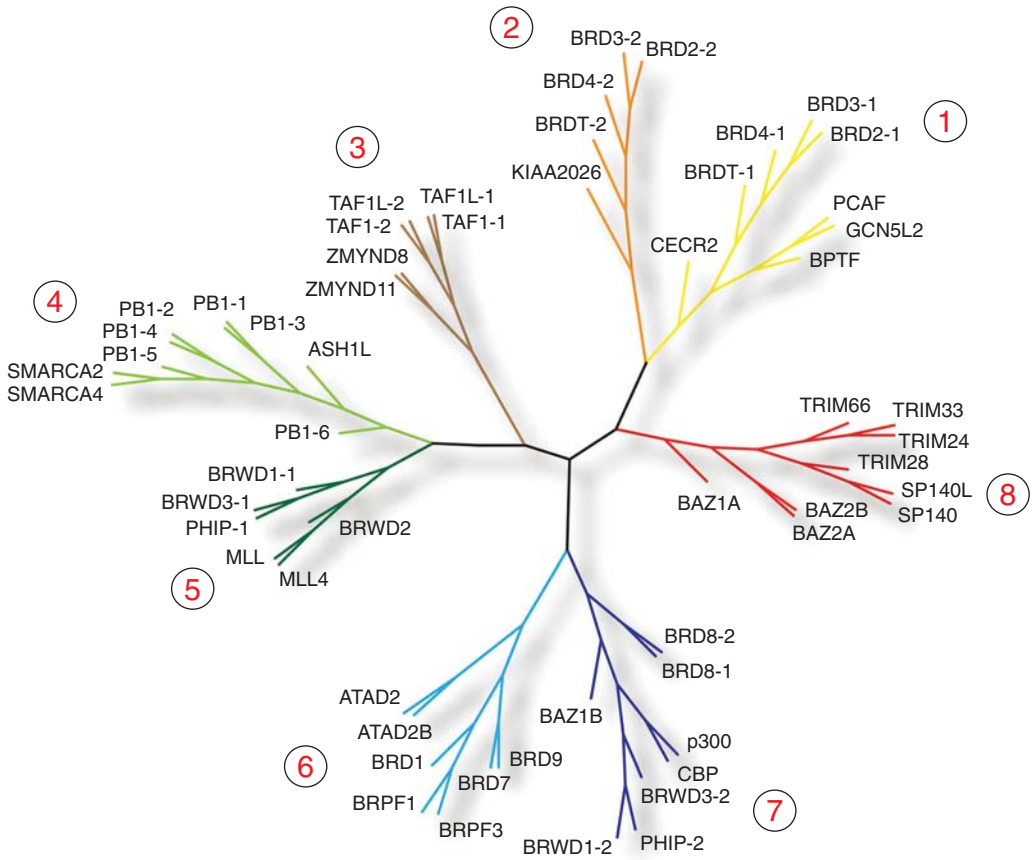


Figure 1.3. The phylogenetic tree of human BET proteins. Adapted from Marmorstein and Zhou[15]. The dendrogram depicts sequence similarity between the 56 human bromodomains. Bromodomains within proteins containing double bromodomains are labeled -1 and -2. Circled numbers represent clustering of bromodomains into discrete families. Cluster 1 and 2 contain the first and second bromodomains of BET proteins, respectively, indicating that the first and second bromodomains across BET proteins are more highly related than the first and second bromodomains within each protein.

Early work indicated that BET proteins BRD2, BRD3 and BRD4 were ubiquitously expressed and provided critical functions in multiple biological processes[19,20]. Mouse embryos depleted of BRD2 or BRD4 are non-viable, indicating that BRD2 and BRD4 are both individually essential for embryonic development[21–23]. A BRD3 knockout mouse has not been reported. BRDT expression is restricted to the testis where it is required for spermatogenesis in mice[19,24]. BRD2 was the first

mammalian BET protein identified. It was initially characterized as a nuclear kinase associated with cell proliferation[25]. However, it had sequence similarity to the Drosophila embryonic development regulator FS(1)H, which when mutated resulted in embryonic lethality[26,27]. It also resembled another mammalian paralog, later identified as BRD4, which associated with the Mediator transcriptional complex[28]. Subsequent characterization of BRD2 revealed that like BRD4, it associated with a vast network of chromatin regulators and transcriptional cofactors[29–31], suggesting that BET proteins were integrated into the transcriptional machinery. The molecular nature and functional significance of many of these interactions will be detailed in later sections. Imaging studies confirmed that BRD2 and BRD4 both localize to chromatin in an acetylation- and bromodomain-dependent manner[32–34]. Their association with chromatin is also mediated at least in part by protein dimerization and/or association with other protein complexes in a bromodomain-independent manner[35,36].

The above studies suggest that BET proteins bind chromatin predominantly via their bromodomains, associate with transcriptional regulators, and are required for proper embryonic development, but what is the evidence that BET proteins directly contribute to transcription? An early piece of evidence suggesting their direct role in transcription was the observation that yeast strains deficient in the BET paralogue BDF1[37] shared phenotypes strongly reminiscent of strains defective in general transcription cofactors[38]. In addition, a gain-of-function reporter system was used to study the Drosophila paralog FS(1)H[39]. In this assay, overexpression of FS(1)H increased reporter activity only when the reporter's promoter was intact. Gel-shifts assays demonstrated direct binding of FS(1)H to the promoter sequence,

indicating that FS(1)H activated the reporter through direct binding to this promoter. More definitive evidence for a role of BET proteins in transcription was obtained using in vitro transcriptional elongation assays. These experiments demonstrated that 1) BRD2, BRD3 and BRD4 were each capable of facilitating transcriptional elongation through chromatin templates, 2) that this activity only occurred on acetylated nucleosome substrates, and 3) intact bromodomains were required[40,41]. In summary, multiple studies provided evidence that BET proteins directly facilitate transcription in an acetylation- and bromodomain-dependent manner.

BET protein structure-function

BET proteins are classified as a family based on the shared and modular domain structure for which they are named. These eponymous domains include the two tandem N-terminal bromodomains (BD1 and BD2) and a C-terminal domain called the extraterminal motif (ET)[18]. The bromodomains associate with acetylated histones on chromatin and other acetylated proteins. The ET domain is a protein-protein interaction module. This stereotypical domain structure (Figure 1.4) thereby allows BET proteins to function as molecular scaffolds onto which multiple transcriptional regulators bind acetylated chromatin. Disruption of this scaffold is the basis for anti-BET therapeutics, as discussed later. In addition to the bromodomains and ET domain, many other conserved features have been identified (Figure 1.4, Table 1.1). The remainder of this section will discuss the structure-function relationship of BET protein domains in greater detail.

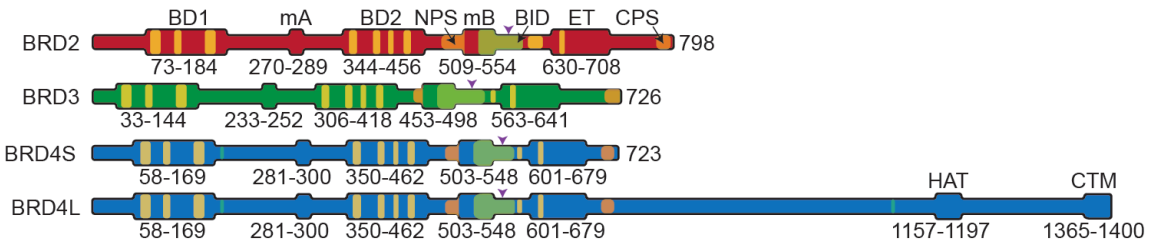


Figure 1.4. Conserved protein domains and other features across the BET family. BET proteins share two tandem bromodomains, BD1 and BD2, which bind to acetylated proteins, as well as the ET protein interaction domain. Additional domains include motif A (uncharacterized) and motif B (a dimerization domain). Two protein isoforms of BRD4 are depicted, a short (BRD4S) and long (BRD4L). The long isoform contains the CTM domain. Known phosphorylated sites are indicated (NPS and CPS). Together, the NPS and BID comprise a phospho-switch. Putative atypical kinase domains are highlighted in orange and the kinase catalytic site is marked with a purple arrow. The putative acetyl-coA binding sites (depicted in teal) and a putative histone acetyl transferase domain (HAT) specific to BRD4L are indicated.

Domain Name	Abbreviation	BRD2	BRD3	BRD4S	BRD4L
Bromodomain 1	BD1	73-184	33-144	58-169	58-169
Motif A	mA	270-289	233-252	281-300	281-300
Bromodomain 2	BD2	344-456	306-418	350-462	350-462
Motif B	mB	509-554	453-498	503-548	503-548
Extraterminal motif	ET	630-708	563-641	601-679	601-679
Histone acetyltransferase domain	HAT	-	-	-	1157-1197
C-terminal motif	CTM	-	-	-	1365-1400
Atypica kinase domain	VIA	80-94	40-54	65-79	65-79
Atypica kinase domain	VIB	112-122	72-82	97-107	97-107
Atypica kinase domain	IX	154-169	114-129	139-154	139-154
Atypica kinase domain	VIA (2)	352-363	314-325	358-369	358-369
Atypica kinase domain	VIB (2)	385-395	347-357	391-401	391-401
Atypica kinase domain	VIII	407-413	369-375	413-419	413-419
Atypica kinase domain	IX	427-437	389-399	433-443	433-443
Atypica kinase domain	I-II	598-619	547-552	584-590	584-590
Atypica kinase domain	III	641-648	574-581	612-619	612-619
N-terminal phosphorylation site	NPS	479-510	441-454	485-504	-
Basic domain	BID	530-591	474-539	524-580	-
C-terminal phosphorylation site	CPS	774-794	705-724	699-717	-
N-terminal acetyl coA binding site	-	-	-	175-180	175-180
C-terminal acetyl coA binding site	-	-	-		1097-1102
Kinase Catalytic Site	-	572	521	563	563

Table 1.1. Amino acid sequences of mouse BET protein domains and features. Sequences are depicted in Figure 1.4.

Evolution and conservation of BET proteins

As noted above, BET proteins share a stereotypical and modular domain structure. Clustering based on bromodomain homology suggests that the BRDT bromodomains are the most ancestral, followed by BRD4 and then BRD2 and BRD3 (Figure 1.3). Clustering of the full-length proteins supports this conclusion by showing that BRDT is the most closely related to FS(1)H/FSH in Drosophila. BRDT then gave rise to BRD4, followed by BRD2 and BRD3 (Figure 1.5). Interestingly, the *Brd2* gene is located within the major histocompatibility complex (MHC) II locus, which is a cluster of genes associated with immune function. The *Brd3* gene is located in a region with MHC-like genes, indicating that it likely spawned from *Brd2* via a gene duplication event[42,43].

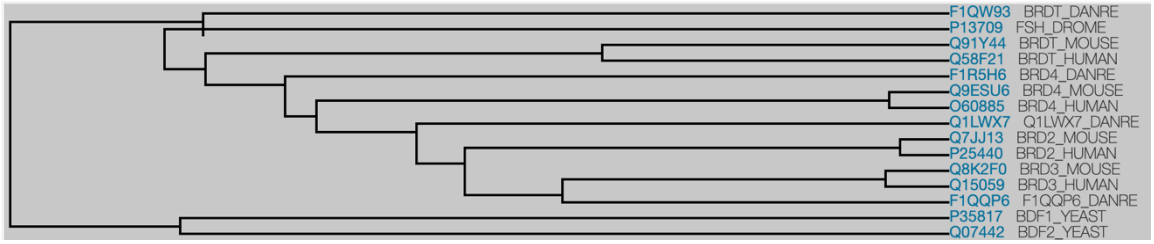


Figure 1.5. Sequence comparison of full-length BET proteins from human, mouse, fish (DANRE), Drosophila (DROME) and yeast. F1QQP6 and Q1LWX7 are the longest isoforms of BRD3 and BRD2 in DANRE, respectively. The dendrogram is the result of sequencing alignment of the indicated proteins in UniProt.

The bromodomains

The BET bromodomains, as with bromodomains in other proteins, are the acetyl-recognition domains[13,44]. Deletion of these domains, or point mutations in their acetyl-lysine recognition pocket, dramatically reduces the ability of BET proteins to associate with chromatin[32–36]. The BET bromodomains are the targets of small molecule inhibitors being tested in various preclinical and clinical settings[45]. The details of these inhibitors are described in more detail below. The

amino acid sequences of the BD1 and BD2 bromodomains across the BET family are ~75% identical, whereas the sequences of BD1 and BD2 within a given protein are only ~44% identical, indicating that BD1 and BD2 are more evolutionarily distinct than the proteins (Figure 1.3)[46]. Nevertheless, the structures of the BET bromodomains reveal remarkable similarity overall[44,46–52]. The bromodomains generally consist of four alpha helixes (α_Z, α_A, α_B, α_C) and two interconnecting loops (ZA and BC)[45]. The four helices form a left-handed bundle around a hydrophobic core with a deep acetyl-recognition cleft formed by juxtaposition of the two intervening ZA and BC loops. The structural elements of the helical bundle bear strong resemblance across the bromodomains. In contrast, the ZA and BC loops comprising the acetyl-lysine binding pocket are more structurally distinct. Though the residues in the ZA and BC loops that make contact with the acetyl-lysine are highly conserved[47,53], sequence divergence among the adjacent residues is thought to contribute to the specificity of each bromodomain for given acetylated substrates.

BET proteins have been shown in many contexts to bind preferentially to multi-acetylated peptides. Because the initial structural characterization of BET bromodomains—in particular those of BRD2[46,49,51,52] and BRD4[47,53]—was performed in the context of mono-acetylated peptides (H4K12ac for BRD2 and H3K14ac for BRD4), it was hypothesized that preference for multi-acetylated substrates was due to the tandem BET bromodomains binding synchronously to individual acetylated lysines of a di-acetylated peptide. However, the co-crystal

structure of BRDT-BD1 with a di-acetylated H4K5ac/8ac peptide revealed an alternative explanation for multi-acetylated substrate binding[48]. The acetyl-binding pocket of BRDT-BD1 can actually accommodate both acetylated lysine residues on its own. The first acetyl-lysine engages with BD1 using the canonical binding cleft described above. The second acetyl-lysine contacts hydrophobic residues on the outside surface of the primary binding pocket. The negative charge imparted by the second acetyl group forms a hydrogen bond with a water molecule that bridges with the first acetyl group, creating a favorable binding surface. Thus, BRDT-BD1 binds preferentially to di-acetylated peptides. Unlike BD1, BD2 only engages with one acetylated lysine due to amino acid differences within BC loop which prevent BD2 from interacting with the second acetylated lysine. Peptide binding assays confirm the distinct binding modalities of BD1 and BD2. BD1 requires at least two acetylated lysines on histone tail peptides for binding to take place, whereas BD2 recognized mono-acetylated peptides with the same affinity as multi-acetylated peptides[48]. Structural homology between BET proteins suggests that these BD1 and BD2-specific binding modalities are likely to be shared across the BET family. Indeed, subsequent solution structures of the BRD3-BD1 in complex with a di-acetylated GATA1 peptide and BRD4-BD1 with various di-acetylated histone tail peptide isoforms (H4K5ac/8ac, H4K12ac/16ac, and H4K12ac/20ac) confirmed this di-acetyl binding mechanism[44,50].

Interestingly, comparison of BRDT bromodomain binding to in vitro reconstituted acetylated nucleosomes versus histone peptides revealed that BD1 binds 6-fold more strongly to acetylated nucleosomes than to acetylated histone

tail peptides[54]. This enhanced affinity for nucleosomes is due to a positively charged patch in close proximity to DNA on the first alpha helix of BD1 (αZ) which binds to DNA directly. This additional bivalent binding modality further diversifies the in vivo substrate binding preferences of the individual BET bromodomains. For example, the DNA binding patch is absent on BRDT-BD2 and though present on BRD3-BD1 and BRD4-BD1, does not appear to bind DNA in those cases[54].

Recognition of diacetyl H4K8ac/12ac is a special case both for BRD4-BD1 and BRD2-BD1. Though BRD4-BD1 binds simultaneously to both of the acetyl groups of many diacetylated peptides, for di-acetylated H4K8ac/12ac, it only recognizes one acetyl at a time, indicating that the exact context of the diacetylated substrate dictates the binding modality used by the bromodomain[44]. BRD2-BD1 also behaves differently in terms of H4K8ac/12ac recognition. Like BRD4-BD1, BRD2-BD1 does not recognize the di-acetylated surface of this particular substrate and instead only associates with one of the acetyl groups. However, unlike BRD4-BD1, which can interact with either of the acetyl groups, BRD2-BD1 only recognizes H4K12ac. In fact, acetylation of H4K8ac interferes with the binding of BRD2-BD1 to H4K12ac. The mechanism for this inhibition is that H4K8ac interferes with the BRD2-BD1 dimerization, which is important for BRD2-BD1 recognition of H4K12ac[46,51]. Dimerization does not appear to occur for any of the other BET bromodomains, making this feature unique to BRD2-BD1. Thus, the sequence context of an acetylated substrate affects whether the acetyl-lysine is recognized by different individual BET bromodomains and the modality by which it

is recognized (mono- versus di-acetyl binding). A large-scale study of this phenomenon revealed that recognition of acetylated histone peptide substrates is influenced not only by the acetyl groups on adjacent residues, but also by methyl- and phospho-marks on these flanking residues and by the length and sequence of the linker region between any two modifications[44].

How do these structural considerations relate to substrate binding preferences of the full-length BET proteins? Histone marks enriched in vivo by each full length BET protein include various combinations of di-, tri-, and tetra-acetylated H4K5/8/12/16 preferentially over monoacetylated H4[55]. Of the di-acetylated H4 peptides, H4K8ac/16ac and H4K5ac/8ac were the most strongly enriched. This was true even in the case of BRD2, which was originally thought to prefer mono-acetylated residues – specifically H4K12ac – based on early but limited histone interaction studies[32] and subsequent structural evaluation of its bromodomains[46,49,51,52]. The direct binding of full-length BET proteins with multi-acetylated histones is supported biochemically by the bromodomain structures described above as well as various histone peptide and nucleosomal binding studies, the latter of which suggest that some BET proteins may associate more tightly with chromatin in vivo via bivalent DNA-histone nucleosomal interactions[54]. Indeed, genome-wide chromatin distributions of BRD2 in vivo correlate better with di-acetylated H4K5ac/8ac, a direct substrate, than with either H3K27ac or H3K4me3, which are histone marks associated with transcription but not predicted to directly bind to BET proteins[56]. Nevertheless, it is likely BET proteins affiliate with large multi-unit complexes which may influence to some extent where BET

proteins are found on chromatin and the chemical makeup of this chromatin. Thus, the specificity and selectivity of BET proteins for substrates in vivo may be determined not simply by the structures of the acetyl-recognition pocket in the bromodomains, but also by protein-protein complex formation that may anchor BET proteins to regions not necessarily enriched for preferred bromodomain substrates. An example of this is the observation that H2A.Z-containing nucleosomes preferentially enrich BRD2 over BRD3 and BRD4[57–59] despite no preference of the BRD2 bromodomains for binding acetylated H2A peptides[44]. A comprehensive comparison of histone binding preferences and chromatin occupancy is needed to better understand to what extent the bromodomains drive genome-wide binding of each BET protein.

BET proteins also interact with acetylated transcription factors. BRD3-BD1 binds to diacetylated GATA1 using a mechanism strikingly similar to acetylated histone recognition (two acetyl-lysines spaced three residues apart)[50]. A similar binding mechanism also occurs between BRD4-BD2 but not BRD4-BD1 and diacetylated TWIST[60], which is interesting because the opposite would be expected based on recognition of diacetylated histone peptides[44,48]. Nevertheless, this arrangement allows BRD4-BD2 to bind acetylated TWIST and BRD4-BD1 to bind acetylated H4, which may be a general mechanism by which BET proteins anchor acetylated transcription factors to chromatin. BRD4 also binds to di-acetylated ERG[61], mono-acetylated RELA (a component of the NF-kB transcription factor complex)[62] and to the acetylated androgen receptor[63]. The binding mechanisms for these interactions have been investigated to varying detail, but

more generally it also remains unclear whether any of the above transcription factor interactions are specific for particular BET proteins. Nevertheless, the ability of BET proteins to recognize acetylated transcription factors may be as functionally significant as their ability to bind to acetylated histones.

The ET domain

The bromodomains are largely responsible for acetyl-lysine recognition and the association of BET proteins with chromatin. However, the ET domain is also critical for BET protein function, and like the bromodomains, is strongly conserved across the BET family. The ET domain has been shown to mediate protein interactions with several transcriptional co-activators including JMJD6, ATAD5, GLTSCR1, CHD4, and NSD3 (WHSC1L1)[31,64–66]. The interaction between BRD4 and JMJD6 was shown to promote enhancer-mediated promoter-proximal pause-release of RNA polymerase II (POL)2[67], and the interaction with NSD3 was shown to promote leukemic gene expression and cell growth through CHD8[65]. The ET domain also functions as the binding site for IN (from murine leukemia virus)[68] and LANA (from Kaposi's sarcoma-associated herpesvirus)[69] to promote viral integration and latency, respectively, and it maintains HIV latency through an association with endogenous BRG1 in HIV-infected cells[70]. The ET domain also associates with CHD4, a component of the NuRD complex which functions in both transcriptional activation and repression[71]. Structural characterization of the ET domain in complex with many of these substrates reveals a common binding motif recognized by the ET domain[31,64,66,68,71,72]. More recently, the association between BRD4 and the cohesin-loading factor NIPBL was shown through yeast two-hybrid

screening assays to be mediated by the ET domain, though it is unclear if NIPBL contains the same ET-binding motif[73,74]. Interestingly, mutations in BRD4 and NIPBL are both associated with Cornelia de Lange syndrome (CdLS), suggesting a causal link. It is possible that the BRD4 mutations disrupt a functional interaction between BRD4 and NIPBL as several of the BRD4 mutations are located in the bromodomains or ET domain[74,75]. This highlights the role of BET proteins as molecular scaffolds onto which additional factors bind to acetylated chromatin. Mutations in any of the critical regions of BRD4 may generate hypomorphic BRD4 alleles unable to effectively recruit transcriptional regulators to chromatin. This same biology is the basis for therapeutic BET bromodomain-disrupting drugs that inhibit disease-promoting pathways by dislodging transcriptional regulators from chromatin. Of note, there is no indication that ET-mediated interactions are specific for one particular BET protein. Thus, as with the bromodomains, it is unclear if the ET domains contribute to BET protein specific functions.

The BRD4L-specific C-terminal tail and the short isoform of BRD4

The *Brd4* gene encodes both a short (BRD4S) and long (BRD4L) isoform. BRD4L is structurally distinct from BRD2, BRD3, and BRD4S in that it possesses an extended tail with a protein-interaction domain located at the C-terminus (Figure 1.4). This C-terminal motif (CTM) binds to polymerase elongation factor beta (PTEFb)[76]. PTEFb is a dimer composed of CCNT1 and CDK9, the latter of which phosphorylates POL2. Recruitment of PTEFb and phosphorylation of POL2 is a key molecular switch that transitions POL2 from its paused configuration to its elongation mode[77]. The CTM of BRD4 is thus thought to recruit PTEFb to gene

promoters to facilitate transcriptional elongation[78]. BRD4L also contains a putative histone acetyltransferase catalytic (HAT) domain and acetyl-CoA binding sites that enable BRD4L to directly acetylate H3K122, which destabilizes nucleosomes and de-compacts chromatin (Figure 1.4)[79]. A screen of mutations that disrupt BRD4 activity validates that in addition to the bromodomains and the ET domain both the CTM and the HAT of BRD4L also contribute to its overall function[80].

BRD4S lacks the extended C-terminal tail of BRD4L due to the presence of a stop codon in an alternatively spliced exon. BRD4S therefore lacks the HAT and CTM domains and more closely resembles BRD2 and BRD3[18]. The regulation of this alternative splicing event has not been studied, and the functional significance of the ratio of BRD4S and BRD4L protein isoforms has not been directly tested, though it may drive certain phenotypes such as tumor growth and metastasis in cancer cells[81]. BRD4S localizes to the nuclear periphery where it interacts with different protein partners in comparison to BRD4L[82]. Importantly these changes in nuclear compartmentalization occur despite BRD4S and BRD4L containing the same acetyl-binding bromodomains, indicating that additional protein interaction domains can drive BET protein localization. BRD4S also functions to maintain HIV latency in infected T cells by recruiting repressive BRG1, a component of a chromatin remodeling complex[70], and like BRD2 and BRD3, can recruit the viral latency protein LANA1 in KSHV-infected cells[69]. However, its expression and function in most cell types remains poorly characterized.

Additional conserved features

In addition to the domains described above, BET proteins also possess a conserved regions motif A and motif B (Figure 1.4). Whereas no function has been ascribed to motif A, limited analysis of motif B indicates that it facilitates homo- and hetero-dimerization of BET proteins and their association with chromatin[35]. Motif B also interacts with LYAR, which contributes to BET occupancy on chromatin and promotes transcription, particularly of ribosomal genes[83,84].

BET proteins possess shared serine-rich patches in the N- and C-terminal regions termed the NPS and CPS respectively (Figure 1.4). Phosphorylation of the NPS induces a conformation change that affects chromatin occupancy[85]. This phospho-switch mechanism was proposed for BRD4, but based on conservation of the relevant residues, may also apply to the other BET proteins. In the unphosphorylated state, BD2 forms an intra-molecular interaction with NPS that prevents BD2 from engaging chromatin. Phosphorylation of the NPS releases BD2 in favor of a basic region termed the BID, allowing BD2 to bind freely to acetylated chromatin. The degree of phosphorylation may thus determine the substrates to which BET proteins bind. Interestingly, hyperphosphorylation of BET proteins may result in bromodomain-independent occupancy on chromatin through gained interactions with the Mediator complex, though it is unclear if this is related to the phospho-switch mechanism above[86]. Thus, the binding of BET proteins to acetylated substrates is not only determined by the residues in contact between the bromodomains and the acetylated peptide substrates, but also by tertiary

folding of the full length BET proteins, which is subject to regulation through post-translational processes such as phosphorylation.

Perturbing BET protein function

BET proteins have received considerable attention as therapeutic targets. This is due in part to their direct involvement in a rare cancer called NUT midline carcinoma (NMC) and their indirect role in more common cancers such as acute myeloid leukemia (AML) among others[87]. However, the large majority of data implicating BET proteins in disease is based on the exploratory use of BET inhibitors in various disease models. It is important to note that BET proteins are critical transcriptional regulators. Thus, any therapeutic benefit derived from BET inhibition may be limited by unwanted effects on normal transcription. For example, it has been shown that global reduction of BRD4 in the adult mouse results in multiple deficits including weight loss, epidermal hyperplasia, and stem cell depletion within the small intestine[88]. Depletion of BRD2 can result in bone marrow failure[89]. This section will therefore summarize the initial observations that drove the development of BET inhibitors. It will then discuss subsequent studies exploring their role in other diseases. And it will conclude with the transcriptional effects associated with BET inhibition. The goal is to provide a lens through which the mechanism of action of these drugs and their potential side effects can be understood.

Role of BET proteins in disease

NUT midline carcinoma is a rare but aggressive squamous cell carcinoma characterized by chromosomal translocations directly involving BRD3 or BRD4[90].

Translocations involving BRD2 have not been described. The translocations result in chimeric proteins comprised of the intact chromatin-binding domains of BRD3/4 and the transcriptional activation domain of the NUT protein. The BRD3/4-NUT fusion protein recruits the histone acetyl transferase EP300 leading to a cycle of hyperacetylation that extends across large swaths of DNA[91,92]. This hyperacetylation in turn leads to aberrant expression of the underlying genes and ultimately cancer[93]. It is unclear why this malignancy only occurs at midline anatomic structures of the head, neck, and mediastinum but in addition to the onset of disease in young adults, this anatomic pattern suggests an embryonic origin. In addition to being direct mediators of NMC, BET proteins were also identified in global candidate screens that looked for susceptibility genes in AML[80,94]. Mechanistically, the dependency of AML cells on BET proteins for cell growth is based on BET proteins driving oncogenic *Myc* expression.

Inhibition using small molecules

Based on this clinical need, a new class of drugs targeting BET proteins was developed. Evolution of candidate compounds and chemical screening approaches converged on thienodiazepine chemistry to target the BET bromodomains[95,96]. Due to the structural similarity of the BET bromodomains, these inhibitors typically do not discriminate between individual BET proteins. The first proof-of-principle use of these pre-clinical compounds was in NMC, where the inhibitor directly targets the translocated BRD3/4-NUT oncogene product[95]. Treatment with the BET inhibitor JQ1 significantly displaced BRD4 and BRD4-NUT from chromatin, induced growth arrest, and differentiated the malignant cells into

a more benign, squamous state. Gene expression analysis indicated that sub-micromolar doses of JQ1 did not simply suppress all genes, but rather selectively turned off the oncogenic program driven by BRD4-NUT[95]. A second thienodiazepine-based BET-specific inhibitor, I-BET, was shown to displace BRD4 from tetra-acetylated H4 peptide, indicating that bromodomains inhibitors outcompete the natural ligands of BET proteins[96]. Strikingly, I-BET prevented LPS-induced inflammatory gene expression in macrophages, which had previously shown to require BRD4[97], while minimally affecting non-inducible genes.

BET inhibitors have been shown to disrupt cell growth in a number of hematological and solid malignancies. An incomplete list of these studies include diffuse large B cell lymphoma[98], multiple myeloma[99], triple negative breast cancer[100], castration-resistance prostate cancer[63], melanoma[58], Ewing sarcoma[101], liposarcoma[102], and glioblastoma[103]. BET inhibitors have been used in other disease settings as well. In cultured pancreatic islet cells, inhibition of BET proteins may protect against the development of insulin resistance[104]. In hepatic cells, targeting BET proteins protects against fibrosis by blocking multiple profibrotic transcriptional pathways[105]. And under cardiac stress conditions, targeting BET proteins prevents cardiomyocyte hypertrophy and cardiac remodeling[106]. Thus, the exploratory use of BET inhibitors has demonstrated that in addition to treating many solid and hematological malignancies, these drugs may have additional use in inflammatory, metabolic, and cardiac disorders.

Based on these promising preliminary findings, several generations of BET inhibitors have since been developed and are at various stages of preclinical and

clinical testing[45,107]. For example, bivalent BET inhibitors using linkers demonstrate stronger anti-leukemic efficacy than molar equivalents of their monovalent counterparts. This enhanced avidity may be due to the ability of individual bromodomains to interact with multiple acetylated lysine residues simultaneously or to the ability of BET proteins to engage two substrates using their tandem bromodomains. Inhibitors that discriminate between the first and second bromodomains (BD1 and BD2) have also been developed[108–110]. Comparison of BD1- and BD2-specific transcriptional effects largely corroborate structural studies indicating that BD1 is the more critical bromodomain for substrate binding. However, the less drastic transcriptional impact caused by BD2 inhibition can be leveraged therapeutically to target more aggressive cancers with better specificity[109] or to prevent particular phenotypes, such as inflammatory gene expression, without altering baseline genes[110]. Interestingly, BD1 inhibition largely recapitulates BD1+BD2 inhibition in terms of antiproliferative effect, whereas BD2 inhibition more directly targets inducible gene expression during inflammation[110]. In summary, BET bromodomain inhibitors block BET protein activity by competitively occupying their acetyl-lysine binding pocket, which displaces the natural acetylated peptide ligands and forces the disassociation of BET proteins from chromatin[30,95,96]. These drugs may have widespread clinical use, but a better understanding of BET protein function is required. Therefore, the transcriptional consequences of removing BET proteins from chromatin will be discussed below.

Transcriptional mechanisms revealed by BET inhibition

There is striking consistency across experimental approaches that perturbing BET protein function reduces transcriptional elongation[41,67,106,111,112]. This is consistent with several lines of evidence implicating BET proteins in promoting POL2 pause-release and/or functioning as histone chaperones to allow POL2 processivity through nucleosomes. BET proteins can complex with PTEFb[30], a known transcriptional elongation factor that releases POL2 from its paused state into its elongating state[77]. BET proteins found at enhancers may contribute to POL2 pause-release by recruiting JMJD6 to enhancers which then form enhancer-promoter contacts that allow JMJD6 to release PTEFb from its inhibitory complex[67]. In support of this model, some studies suggest that degree of BET protein enrichment at enhancers correlates with BET inhibitor sensitivity[98,99]. However, these PTEFb-related mechanisms may not be as functional as originally thought, given that BET proteins can affect elongation independent of PTEFb occupancy[111]. In addition to pause-release, assays using highly purified components have shown that BET proteins can each individually stimulate transcriptional elongation independent of the PTEFb-interacting CTM domain when added to in vitro transcription reactions and that this activity is sensitive to bromodomain inhibition[40,41]. It was therefore proposed that BET proteins are histone chaperones that assemble and disassemble nucleosomes from acetylated histones to allow the passage of POL2. Though the biochemical basis for this hypothesis remains unexplored, this model is supported by the fact that the amount

of BET protein enriched on gene bodies correlates better with transcriptional activity than the amount enriched at gene promoters[41].

In addition to pause-release and elongation, BET proteins also function in transcriptional initiation through their interactions with acetylated transcription factors. Displacement of BET proteins from chromatin via BET bromodomain inhibition has been shown to decrease transcription factor binding resulting in concomitant changes in gene expression[61–63,96,113]. Our lab has shown that in the case of GATA1, BET proteins may play a role not only in GATA1 recruitment and thus transcriptional initiation, but also in downstream transcriptional events that take place after GATA1 binding has been established[113]. The role of BET proteins in establish transcription factor occupancy may be key to the observation that "inducible" genes tend to be the most sensitive to BET inhibition[96]. This indicates that lineage- or state-dependent transcriptional programs – whether oncogenic, inflammatory, or stress-induced – may be more easily targeted by therapeutic BET inhibitors.

Predicting genes that are vulnerable to BET inhibition remains a challenge. No single chromatin attribute, including proximity to enhancers or degree of BET enrichment, adequately predicts a gene's primary responsiveness to BET inhibition[112]. Moreover, comparison of JQ1-sensitive and JQ1-insensitive AML cell lines have shown that a core set of genes – including oncogenic *Myc* – are hypersensitive to BET inhibition regardless of the resultant effect on cell growth, indicating that even if a gene's sensitivity could be predicted, the ultimate effect on cell phenotype cannot be predicted[112]. This suggests that for some transcriptional

programs, the cellular response to BET inhibition does not depend on the primary transcriptional response but instead on secondary cell-specific attributes. Plasticity within these transcriptional pathways may also be at fault for the development of resistance to BET inhibitors over time[86,114–116]. Therefore, a better understanding of the exact role BET proteins play in transcription is needed to guide the development of more dependable anti-BET therapeutic strategies.

In scenarios where BET inhibition has a favorable outcome, the therapeutic efficacy is often attributed to BRD4 inhibition. This is in part due to the proposed BRD4-PTEFb molecular mechanism considered to be important for transcriptional pause-release and elongation[77]. Though many studies do not investigate the independent roles of BRD2/3/4, in the case of AML, depletion of BRD2 or BRD3 does not affect cell proliferation or POL2 phosphorylation whereas depletion of BRD4 does[80,94,112]. However, in other cases, BRD2 and BRD3 do in fact contribute, which will be discussed below. Thus, though BRD4 certainly contributes to transcription, it is important to remember that BET bromodomain inhibitors typically used in these studies do not discriminate between individual BET proteins. Their therapeutic effect is thus the aggregate result of inhibiting the entire BET family. Identifying the individual roles of BET proteins in transcription is critical to understanding the mechanism of action of BET inhibitors.

Evidence for distinct BET protein functions

BET proteins are epigenetic reader proteins – adaptor proteins with specialized domains that bind to acetylated chromatin and recruit multi-protein complexes. However, it remains unclear whether individual BET proteins function

distinctly or redundantly. Identifying distinct functions for individual BET proteins is complicated by their propensity to dimerize with one another[35], to associate with similar multi-protein complexes[30,31,64], and to co-localize on chromatin[63,102,117,118]. Given these findings, one can imagine models where BET proteins function interdependently within multi-unit complexes. The purpose of this section is to synthesize the evidence that individual BET proteins function non-redundantly and thus possess specialized functional niches on chromatin.

Insights revealed by genetic studies

The phenotypes of genetic knockout mice provide the strongest evidence that BET proteins function non-redundantly. For example, both BRD2 and BRD4 are independently required for embryonic development, and the phenotypes of these mice are distinct. BRD4-defects manifest earlier during development resulting in early post-implantation defects, whereas BRD2-defects occur later during development resulting in impaired neural tube closure[21–23]. Interestingly, mice with hypomorphic *Brd2* alleles (resulting in global reduction of BRD2 protein across tissues) are viable but display metabolic irregularities, indicating that even the levels of BRD2 are critical[119]. BRD3 knockout mice have not been reported.

The requirement for individual BET proteins can also be identified through targeted depletion strategies. For example, depletion of BRD2, BRD3 or BRD4 each reduces proliferation of castration-resistant prostate cancer cells and diffuse large B cell lymphoma cells[63,98]. Likewise, depletion of all three BET proteins effects cell growth in Ewing sarcoma cells[101,120,121]. Both BRD2 and BRD4 are individually required for the differentiation of CD4+ Th17 cells[122], and all three BET

proteins contribute to LPS-induced stimulation of cytokine expression in macrophages[123]. These studies indicate that like BRD4, BRD2 and BRD3 are also required for many biological processes.

Global transcriptional signatures and phenotypes are also different depending on whether BRD2, BRD3 or BRD4 are depleted. For example, in pancreatic beta cells, depletion of either BRD2 or BRD4 enhances transcription of insulin, whereas inhibition of BRD2 but not BRD4 enhances fatty acid oxidation[104]. A global analysis of gene expression changes in these cells revealed that depletion of BRD2, BRD3 and BRD4 each yielded different signatures, indicating that individual BET proteins are involved in both shared and discrete transcriptional pathways[104]. A similar gene expression analysis performed in breast cancer cells revealed distinct effects on a panel of epithelial-to-mesenchymal transition (EMT) genes upon depletion of BRD2, BRD3 or BRD4. Phenotypically, BRD2 appeared to promote EMT whereas BRD3 and BRD4 repressed EMT[124]. Likewise, in skeletal myogenesis, depletion of BRD3 and BRD4 have opposing effects on myogenic differentiation with BRD4 appearing to promote differentiation and BRD3 appearing to inhibit it (BRD2 depletion was not achieved so was not examined further)[125]. In embryonic stem cell differentiation into myoblasts, Nodal gene expression is mediated first by BRD4 and then by BRD2 with a concomitant switch in the BET protein predominantly found at cis-regulatory elements near the Nodal gene[126]. This indicates that BRD2 and BRD4 promote Nodal gene expression independently. In blocking interferon-induced gene expression, targeted BRD4-depletion was not as effective as pan-BET bromodomain inhibition, supporting

contributory or perhaps independent roles for BRD2 and BRD3 in transcription[110]. BRD2 and BRD4 also have unique roles in the DNA damage response with BRD2 directly contributing to DNA repair at double-strand breaks[127] and BRD4 functioning as a more general insulator of acetylated chromatin[128]. Collectively, these studies strongly suggest that BRD2, BRD3 and BRD4 function within distinct niches during transcription and other nuclear processes.

Insights revealed by chromatin occupancy studies

The distribution of individual BET proteins on chromatin are largely similar but with some distinctions. A list of studies that perform chromatin immunoprecipitation coupled to high-throughput sequencing (ChIPseq) for two or more BET proteins in parallel is presented as Table 1.2.

Study	Cells/tissue	PMID	BET proteins	Main findings
Nicodeme et al, *Nature* (2010)	Mouse bone marrow derived macrophages	PMID: 21068722	BRD2/3/4	Gene tracks show similarities and distinctions.
LeRoy et al, *Genome Biol* (2012)	Human 293T cells	PMID: 22897906	BRD2/3/4 (Flag-tagged)	Similar metaprofiles across genes
Asangani et al, *Nature* (2014)	Human prostate cancer cells	PMID: 24759320	BRD2/3/4	Colocalization at many regions as well as independent peaks.
Anders et al, *Nature Biotech* (2014)	Human MM1.S multiple myeloma cells	PMID: 24336317	BRD2/3/4	Strong colocalization of BRD2/3/4 at promoters with BRD2/4 but little BRD3 at enhancers
Di Micco et al ,*Cell Reports* (2014)	Human ESC	PMID: 25263550	BRD2/3/4	No comparative analysis
Fong et al, *Nature* (2015)	Mouse MLL–AF9 AML cells	PMID: 26367796	BRD2/3/4	No comparative analysis
Ding et al, *PNAS* (2015)	Human LX-2 cells (hepatic stellate cells)	PMID: 26644586	BRD2/3/4	No comparative analysis
Keck et al, *J Biol Chem* (2017)	Human Mutul cells (EBV-positive Burkitt lymphoma)	PMID: 28588024	BRD2/3/4	Strong colocalization at the EBV lytic origins of replication
Stonestrom et al, *Blood* (2015) & Hsu et al, *Mol Cell* (2017)	Mouse G1E-ER4 erythroblast cells	PMID: 25696920 & PMID: 28388437	BRD2/3/4	Colocalization of BRD3/4 at GATA1 and BRD2/3 at CTCF
Cheung et al, *Mol Cell* (2017)	Mouse CD4+ T cells (Th17 cells)	PMID: 28262505	BRD2/4	Colocalization with more BRD2 at CTCF
Piunti et al, *Nat Med* (2017)	Human SF8628 cells (diffuse intrinsic pontine glioma)	PMID: 28263307	BRD2/4	Colocalization of BRD2/4
Fontanals-Cirera et al, *Mol Cell* 2018	Human melanoma cells	PMID: 29149598	BRD2/4	Similar metagene profiles at promoters and enhancers, perhaps more BRD2 at promoters
Xu et al, *PNAS* (2018)	Human U87 gioblastoma cells	PMID: 29764999	BRD2/3/4	Colocalization at E2F1 target genes
Chen et al, *Nature Commun* (2019)	Human liposarcoma cell lines	PMID: 30903020	BRD2/3/4	Colocalization with more BRD2/3 at promoters and more BRD4 at enhancers
Khoueiry et al, *Epigenetics Chromatin* (2019)	Human K562 cells	PMID: 31266503	BRD2/3/4	Similar metagene profiles at promoters
Bevill et al, *Mol Cancer Res* (2019)	Human breast cancer cells (MDA-MB-231)	PMID: 31000582	BRD2/4	Similar metagene profiles at promoters
Federation et al, *Cell Reports* (2020)	Human K562 cells	PMID: 32101728	BRD2/3/4	Similar gene track profiles
Gilan et al, *Science* (2020)	Human THP-1 cells	PMID: 32193360	BRD2/3/4	Similar gene track profiles

Table 1.2. Studies comparing chromatin distribution of two or more BET proteins. PubMed IDs (PMID) listed for reference.

BET proteins tend be highly enriched at promoters and enhancers, consistent with their role in regulating transcription. There is some indication that BRD2 and BRD3 are slightly more enriched at promoters in comparison to enhancers, whereas BRD4 is more enriched at enhancers[102,117,129]. This is consistent with BRD2 and BRD3 binding more strongly to the promoter-associated histone variant H2A.Z[57–59]. Interestingly, whereas BD1-specific bromodomains inhibitors displace BRD2, BRD3, and BRD4 relatively equally from chromatin, BD2-specific inhibitors affect BRD2 and BRD3 while leaving BRD4 occupancy largely intact[110]. This indicates that BRD2 and BRD3 may bind chromatin using mechanisms inherently differently than BRD4. Nevertheless, all three BET proteins bind promoters where they exhibit similar binding profiles[55,118]. BET proteins may also bind within gene bodies potentially reflecting a role in transcriptional elongation[40,41,55]. However, in comparison to promoter-proximal binding, the degree BET protein binding over gene bodies is not as defined as that of POL2 or other components of the elongating POL2 complex[130]. This indicates that BET proteins may not travel as part of the elongating POL2 complex directly, but the exact details of BET protein function over gene bodies remains unclear. Consistent with the observation that the transcriptional elongation promoting activity is shared by all three BET proteins, no obvious differences in their distributions at promoters or within genes can be appreciated[40,41,55]. In addition to enrichment of BET proteins at promoters, enhancers and gene bodies, we and others have noted some degree of overlap between BRD2 (and BRD3) and the architectural protein CTCF[122,129,131]. This suggests a role for BRD2 and possibly BRD3 in higher order chromatin

organization via the maintenance of topologically associating domains (TADs) and/or transcriptional boundaries. However, the degree of colocalization and the specificity of overlap between CTCF and BRD2 versus BRD4 varies between study and technical replicate. In summary, BET proteins tend to bind similar regions of the genome. Though some distinctions have been identified, a thorough meta-analysis of publicly available data is required to confidently state which differences are robust to experimental parameters and thus reflect true functional specificity.

Insights revealed by protein interaction studies

BET proteins may carry out distinct functions or bind unique regions of the genome by associating with different protein complexes. However, as with the chromatin occupancy profiles described above, the interactomes of BRD2, BRD3 and BRD4 are also largely similar[30,31]. This may be due to the tendency of BET proteins to associate with one another through direct dimerization, their mutual association with active chromatin, or their shared possession of the highly conserved ET protein-protein interaction domain. Nevertheless, one clear difference that is observed is that BRD4 tends to associate more strongly with the PTEFb than does BRD2 or BRD3[30,31]. BRDT also binds to PTEFb[31], which is consistent with both BRD4 and BRDT containing the PTEFb-interacting CTM domain on their extended tail. This is in contrast to interactions with casein kinase II (CK2) which are predominated by BRD2, BRD3 and BRDT[31] and to interactions with the RNA polymerase II associated factor (PAF) complex which occur to some degree between all BET proteins[30]. Immunoprecipitation of ectopically expressed BRD3 was shown to enrich PTEFb, PAF, and CK2 among many other chromatin-

associated proteins[71]. Unlike for PTEFb, BET protein domains that interact with PAF and CK2 are not known, though CK2 has been shown to phosphorylate regions of BRD4 that are conserved with the other BET proteins[85]. Several additional interactions known to be mediated by the ET domain (JMJD6, ATAD5, WHSC1L1/NSD3) are also found in interactome datasets.

In summary, BET proteins are highly enriched at active chromatin where they bind acetylated residues and interact with multiple chromatin regulators, but the mechanisms underlying BET-specific functions in transcription remain unresolved. It has been proposed that BET-specific functions are conveyed in part via selectivity of the bromodomains for distinct acetylated substrates[44,55]. Specificity may also be mediated in a bromodomain-independent manner by protein-interaction domains such as the ET domain that influence co-factor recruitment and perhaps chromatin binding[30,31]. Nevertheless, the chromatin distributions and protein interaction networks of BET proteins are largely similar though with some exceptions. Additional work is required to determine whether these exceptions reflect biologically significant differences in BET protein function.

Functions of BET proteins in GATA1-mediated transcription

Our laboratory has used GATA1-mediated erythroid cell differentiation as a model system to further dissect the functions of BET proteins in transcription[132]. Interestingly, anemia is one of the predominant side effects observed in patients treated with clinical BET inhibitors[133–135]. Thrombocytopenia (reduced platelet count) is another common toxicity. GATA1 drives the differentiation of both erythrocytes and megakaryocytes (the cells that give rise to platelets) from a

common precursor[136]. This clinical picture is highly indicative of deficient GATA1 activity in patients treated with anti-BET therapy. GATA1-mediate transcription in erythroid cells is thus a physiological and clinically relevant model for the study of BET protein function.

GATA1 is acetylated by the acetyltransferases EP300 and CBP[137,138]. The BET protein bromodomain BD1 binds to a di-acetylated peptide within GATA1 using a mechanism reminiscent of acetylated histone peptide recognition[50]. Though this was demonstrated in structural detail for BRD3-BD1, acetylated GATA1 peptides also enrich BRD4[139]. Mutations of the acetylated GATA1 residues recognized by BD1 prevent erythroid maturation and reduce occupancy of GATA1 on chromatin, indicating that GATA1 acetylation is critical for its function in vivo[140]. GATA1-induced gene expression is also drastically reduced by BET bromodomain inhibition, indicating that BET proteins are critical mediators of acetylated GATA1 activity[113,139]. In support of this model, BET proteins colocalize with GATA1 genome wide, are required for GATA1 recruitment, and mediate transcription downstream of GATA1 binding[113]. Using GATA1-mediated erythroid maturation, our laboratory has demonstrated that BRD2 and BRD4 are both uniquely required for GATA1 activity (Figure 1.6)[113]. BRD3 appears dispensable, potentially due to compensation by BRD2[113]. Thus, BRD2 and BRD3 may function independently of BRD4 in these cells.

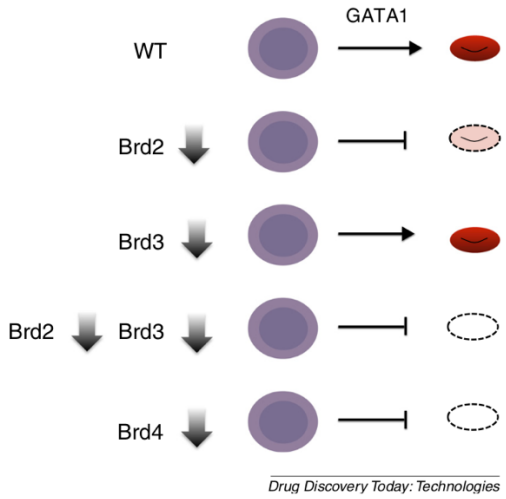


Figure 1.6. BET protein requirements in erythropoiesis. Adapted from Stonestrom et al[132]. BRD2 depletion partially inhibits erythroid maturation. BRD3 depletion has no apparent effect itself, but when combined with BRD2-deficiency, completely blocks maturation. BRD4 depletion prevents maturation on its own.

The distinct functions of BRD2 and BRD4 may be reflected by their unique distribution patterns on chromatin. Whereas BRD4 tightly colocalizes with GATA1 and binds chromatin in a GATA1-dependent manner, BRD2 exhibits a more GATA1-independent occupancy pattern and is also enriched at CTCF-sites[113,131]. It remains unclear whether these differences in chromatin occupancy patterns reflect the distinct functions of BRD2 and BRD4. Moreover, the structural basis for these distinct patterns remain unknown. Identifying the BET protein structural domains that underlie the shared and distinct functions of BET proteins may offer a window of insight into the mechanisms by which individual BET proteins function distinctly or cooperatively in transcription. Moreover, identification of structural features that differ between BRD2 and BRD4 could be exploited to design BET inhibitors that target specific BET proteins. The goal of my dissertation is thus to

use GATA1-mediated differentiation to map the domains that distinguish BET proteins functionally in order to identified BET-specific transcriptional mechanisms.

CHAPTER 2: MATERIALS AND METHODS

Cell lines, cell culture, and GATA1-ER activation

G1E cells are erythroblasts derived from *Gata1*-null male mouse embryonic cells[141]. G1E-ER4 cells are subcloned from G1E cells after expressing a retrovirally inserted construct that encodes GATA1 fused to the ligand binding domain of the estrogen receptor (GATA1-ER), which allows for conditional activation of GATA1 upon administration of estradiol[142]. *Brd2*-null G1E-ER4 cells (BRD2 KO1-KO4) cells were derived by single cell cloning after nucleofection with vectors expressing spCas9 and a gRNA targeting the first bromodomain of BRD2 as previously described[113]. Clones KO1 and KO2 were previously described[131]. G1E-ER4 based cells were grown in suspension flasks in a humidified incubator at 37C at 5% carbon dioxide in G1E media. G1E media consists of the following reagents: Iscove's Modified Dulbecco's Medium (Corning, 10-016-CV); 15% FBS (Gemini, 100-106), 2% Pen Strep (Gibco, 15140-122), 140μM 1-Thioglycerol, (Sigma, M6145), 2U/mL Epoetin (Amgen) and conditioned medium (titered for optimal cell growth) from a kit ligand producing CHO cell line. Cells were maintained in culture at a cell density of 0.1-1.0e6 cells/mL. Erythroid maturation experiments were performed by plating G1E-ER4 based cells at a cell density of 0.4-0.5e6 cells/mL in fresh G1E media containing 100nM estradiol (Sigma, E2758) for 24 hours. For the visualization of hemoglobin production, 40e6 cells were plated as above and matured for 48 hours with estradiol. 293T cells used for retrovirus production were grown on 100mM plasma-treated polystyrene plates in 293T media: Dulbecco's Modified Eagle Medium (Gibco, 11965084), 10% FBS (Gemini, 100-106), 2% Pen

Strep (Gibco, 15140-122), 1% L:-Glutamine 200mM (Gibco, 25030-081) and 1% 100mM sterile-filtered sodium pyruvate. 293T cells were split every 2-3 days using 0.05% Trypsin-EDTA (Gibco, 25300-054).

RNA isolation and RTqPCR

For gene expression analysis, up to 1 million cells were resuspended in 1mL of TRIzol (Ambion, 15596018). 200µl of chloroform (Fisher, BP1145-1) was added to each sample. Samples were vigorously inverted for 15 seconds, rested for 3 minutes and spun at 10,000 rpm for 15 minutes at 4C in a microcentrifuge. The aqueous phase (top layer) was applied to purification spin columns from the RNeasy® Mini Kit (Qiagen, 74106). Subsequent purification steps, including an on-column DNase treatment step (Qiagen, 79256), were followed according to the manufacturer's instructions. RNA was eluted from the column in 40µl RNAse-free water from the kit, and the RNA concentration and purity were quantified using a spectrophotometer (Thermo Scientific, Nanodrop 2000). Reverse transcription was performed on 1µg of RNA using the iScript Reverse Transcription Supermix for RTqPCR (Bio-Rad, 1708841). The cDNA samples were diluted 5x in sterile 0.2µM filtered MiliQ water. RTqPCR was performed using 2µl diluted cDNA,1µl of 9µM forward/reverse primers and 5µl Power SYBR™ Green PCR Master Mix (Thermo Fisher Scientific, 4367660) on an ABI Vii7 real-time PCR machine. qPCR primers were designed to target mature transcripts. RTqPCR primer sequences are found in Table 2.1.

Name	Forward	Reverse	Note
Brd2 5'	TTAAACACCCCATGGACCTC	ATCAGCAGCAAACTCCTGTG	mouse RTqPCR
Brd2 3'	GCTCAACTCCACCAAAAAGC	AGAAGAGGAAGATGACGACGAG	mouse RTqPCR
Brd3	GGAGATGCTGTCCAAGAAGC	TGTGCATCTGGGTACTCTCG	mouse RTqPCR
Brd4 5'	TAAAGTGCTGCAGTGGCATC	TGGAGAACATCAATCGGACA	mouse RTqPCR
Gata1-ER	TATGGCAAGACGGCACTCTAC	TGTTGTTGCTCTTCCCTTCC	mouse RTqPCR
GFP/YFP	AAGGGCATCGACTTCAAGG	ATGCCGTTCTTCTGCTTGTC	mouse RTqPCR
Hbb-b1	AACGATGGCCTGAATCACTTG	AGCCTGAAGTTCTCAGGATCC	mouse RTqPCR
Hba-a1	GTGGATCCCGTCAACTTCAAG	CAAGGTCACCAGCAGGCAGT	mouse RTqPCR
Alas2	TATGTGCAGGCCATCAACTACC	TTTCCATCATCTGAGGGCTGTGG	mouse RTqPCR
Slc4a1	TGGAGGCCTGATCCGTGATA	AGCGCATCGGTGATGTCA	mouse RTqPCR
Spta1	AAAGAGTTCCGCTCTTGCCTGA	TTTCCTCCCTGGATCCACAGCAT	mouse RTqPCR
Csf2rb	AAGAGCCTGCAACTCACTGG	AGGGCTTGTGATTCTTCCCT	mouse RTqPCR
Csf2rb2	GAGCGAGTGGAGCAATGAGTA	ATAAACACGGCCAAAGTGGA	mouse RTqPCR
Kcnn4	TCTCTGGCTCACCACAGCTT	GGAATGTGATCGGAATCAGC	mouse RTqPCR
Rhag	TTCTGGAAATTGCTGTATTTGCT	CTCCAAAGGCATGGATTGTC	mouse RTqPCR
Actb	ACACCCGCCACCAGTTC	TACAGCCCGGGGAGCAT	mouse RTqPCR
Gapdh	AGGTTGTCTCCTGCGACTTCA	CCAGGAAATGAGCTTGACAAAG	mouse RTqPCR
Pabpc1	CTACCAGCCAGCACCTCCTT	TGCAGCACGGTTCTGAGTCT	mouse RTqPCR

Table 2.1. RTqPCR primer sequences used in this study

Relative mRNA levels were quantified using the delta cycle of threshold (Ct) method (Relative mRNA Level = 2^($Ct_{control\ gene}$ – $Ct_{target\ gene}$). All gene expression analyses included at least two of the following housekeeping genes as controls for normalization: *Pabpc1*, *Actb* and *Gapdh*. All RTqPCR experiments were performed at least in triplicate (three independent GATA1-ER induction replicates). Plots illustrating gene expression were generated in RStudio using ggplot2.

RNA sequencing

For the RNA sequencing experiment, parental G1E-ER4, BRD2 KO1 and BRD2 KO2 cells were matured for 24 hours with estradiol prior to harvesting RNA. The experiment was performed in three replicates yielding a total of 9 samples.

Library preparation and sequencing

RNA concentration was determined using the Agilent RNA 6000 Nano kit (Agilent, 5067-1511) and 4µg of total RNA was enriched for mRNA by polyA

selection with Dynabeads mRNA DIRECT Micro Kit (Life Technologies, 61021) using the mRNA Isolation from Purified Total RNA protocol. Concentration of mRNA enriched samples was measured using the Agilent RNA 6000 Pico kit (Agilent, 5067-1513) before proceeding to library preparation with the remaining sample volume using the Ion Total RNA-Seq Kit c2 and IonXpress RNA-Seq Barcode 1-16 kit (Life Technologies, 4475936, 4475485). Final libraries were assessed using the Agilent High Sensitivity DNA kit (Agilent, 5067-4626) before pooling on an equimolar basis as determined by Bioanalyzer (Agilent) molarity estimate of each library. The pool was subsequently sequenced three times on an Ion Torrent Proton sequencer using Ion PI Hi-Q OT2 200 Kit, Ion PI Hi-Q Sequencing 200 Kit and Ion PI Chip Kit v3 (Life Technologies, A26434, A26433, A26771).

Computational analysis

Primary analysis of sequence data was performed by Ion Torrent Torrent Suite software (v4.4.3) before being exported into FASTQ format. Mapping to the mm10 genome (GCF_000001635.24_GRCm38.p4_genomic.gff) was performed using the following STAR options: runThreadN 8, readFilesCommand zcat, outSAMtype SAM,

chimSegmentMin 32, outFilterType BySJout, outFilterMultimapNmax 20, alignSJoverhangMin 8, alignSJDBoverhangMin 1, outFilterMismatchNmax 999, outFilterMismatchNoverLmax 0.04, alignIntronMin 20, alignIntronMax 1000000, alignMatesGapMax 1000000. The following analyses were performed using the indicated R packages/functions: PCA (stats/prcomp), differential gene expression

(DEGandMore/DeWrapper), heatmap visualization (stats/pheatmap). The top ten gene sets enriched in the list of differentially expressed genes were identified from the Molecular Signatures Database hallmark gene sets using the GSEA online tool (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp) with FDR cutoff of 0.05. Plots illustrating gene expression and gene set enrichment were generated with custom code in RStudio using ggplot2. For the comparison with the previous microarray data set assessing the impact of JQ1 treatment on GATA1-ER activity, we downloaded publicly available data from Gene Expression Omnibus (GSE62709: samples G1E GATA-ER +E2 and G1E GATA-ER +E2 +JQ1 replicates 1-3). Genes were included in the merged data if they had an average microarray measurement of 6 or higher in at least one of the groups and had an average normalized RNA-seq read count of 6 or higher in at least one of the groups. A correlation between the two datasets was tested by calculating a Pearson correlation coefficient and a linear regression line was computed. Differentially expressed genes were identified from the microarray based on an average fold change > 2 and FDR < 0.01, which were then intersected with differentially expressed genes from the RNAseq experiment (using same criteria) as depicted by the Venn diagrams in Figure 3.2.G. Statistical significance for each overlap was calculated using a Fisher's Exact Test. For differential isoform analysis, the Ensembl gene mode of mouse *Brd4* that includes 7 known alternative transcripts was used. All reads aligning to *Brd4* by STAR (run with a 2-pass gap alignment mode) were included. The Bioconductor package was used to tabulate the sequencing depth at each *Brd4* exon and the frequency of reads mapping to

junction sites. The reads all three replicates in each group were summed and DEseq2 was used to calculate statistical significances between groups.

Plasmids and cloning

All cloning was performed using standard molecular cloning techniques. PCR was performed using Platinum® *Pfx* DNA Polymerase (Invitrogen, 11708). All constructs were cloned into the MigR1 murine retroviral vector (with an IRES-GFP selection marker) or a PK1-based N-terminal YFP fusion vector. For constructs requiring assembly of multiple PCR fragments, Gibson Assembly® Master Mix was used (NEB, E2611L). The vectors and inserts for each construct are provided in Table 2.2.

Plasmid Name	Vector	Insert
HA-Brd2	MigR1	Brd2(1-798)
HA-Brd3	MigR1	Brd3(1-726)
HA-Brd4S	MigR1	Brd4S(1-723)
HA-Brd4L	MigR1	Brd4L(1-1400)
HA-Brd2-4S	MigR1	Fusion of Brd2(1-456)-Brd4S(463-723)
HA-Brd4-2	MigR1	Fusion of Brd4S(1-462)-Brd2(457-798)
HA-Brd2dET	MigR1	Brd2 with Brd2(629-716) deleted
HA-Brd2dCC	MigR1	Brd2 with Brd2(717-762) deleted
YFP-Brd2	PK1	Brd2(1-798)
YFP-Brd3	PK1	Brd3(1-726)
YFP-Brd4S	PK1	Brd4S(1-723)
YFP-Brd2-4S	PK1	Fusion of Brd2(1-456)-Brd4S(463-723)
YFP-Brd4-2	PK1	Fusion of Brd4S(1-462)-Brd2(457-798)
YFP-Brd2-4mid	PK1	Fusion of Brd2(1-456)-Brd4S(463-599)-Brd2(629-798)
YFP-Brd2-4ET	PK1	Fusion of Brd2(1-628)-Brd4S(600-687)-Brd2(717-798)
YFP-Brd2-4end	PK1	Fusion of Brd2(1-716)-Brd4S(688-723)
YFP-Brd4S-2ET	PK1	Fusion of Brd4S(1-599)-Brd2(629-716)-Brd4S(688-723)
YFP-Brd4S-2CC	PK1	Fusion of Brd4S(1-687)-Brd2(717-762)-Brd4S(688-723)
YFP-Brd4S-2ETCC	PK1	Fusion of Brd4S(1-599)-Brd2(629-762)-Brd4S(688-723)
YFP-Brd2dET	PK1	Brd2 with Brd2(629-716) deleted
YFP-Brd2dCC	PK1	Brd2 with Brd2(717-762) deleted
GST-BRD2-mB	pGEX-2T	Brd2(509-554)
GST-BRD2-ET	pGEX-2T	Brd2(629-716)
GST-BRD2-CC	pGEX-2T	Brd2(717-762)
GST-BRD2-CCp	pGEX-2T	Brd2(717-762, E725P+R732P+D740P)
GST-BRD2-CCd	pGEX-2T	Brd2(717-762, L726D+L734D+L738D)
GST-BRD2-ETCC	pGEX-2T	Brd2(629-762)
GST-BRD3-ET	pGEX-2T	Brd3(562-651)
GST-BRD3-ETCC	pGEX-2T	Brd3(562-694)
GST-BRDT-ET	pGEX-2T	BrdT(495-580)
GST-BRDT-ETCC	pGEX-2T	BrdT(495-624)

Table 2.2. Mouse BET protein expression constructs used in this study.

Cell growth assay

Differences in cell growth upon transduction with retroviruses expressing BET protein constructs were performed using viral titers that yield 30-40% GFP+ cells. GFP% was measured by flow cytometry and tracked over time while

maintaining the cells at a cell density of 0.1-1.0e6 cells/mL. %GFP was normalized by subtracting the starting %GFP (either day 1 or day 2). All cell growth assays were performed in triplicate (three independent transduction experiments).

Retrovirus production and transduction

Retroviruses were produced using calcium phosphate co-transfection of 15μg construct and 15μg of pCL-Eco packaging plasmid into 293T cells. 293T cell media was replaced with G1E media after 24 hours. Virus-containing G1E media was then harvested after an additional 24 hours and passed through a 0.2μm filter. Viral supernatants were diluted in G1E media to find titers that yield 30-40% GFP/YFP+ cells. Retroviral transduction was performed by adding 250μl of titered virus to one million G1E-ER4 (or BRD2 KO) cells in 1mL G1E media supplemented with 4μl/mL polybrene and 10μl/mL 1M HEPES (Gibco, 15630080) in 12-well plates and spinning at 2000rpm in an centrifuge for 90 minutes. The supernatant was removed, and cells were resuspended in 5mL fresh G1E media. Fluorescence was measured 24 hours later on a flow cytometer.

TER119 staining, flow cytometry, and cell sorting

Flow cytometry was performed using standard antibody staining protocols and cytometry settings. Briefly, 10^5 to 10^6 cells were harvested and washed with PBS. For TER119 expression experiments, cells were infected with low-titer YFP-BET encoding (or empty vector control) virus to guarantee single copy gene integration. After 24 hours of rest, estradiol was added to the mixed population of infected/uninfected cells for an additional 24 hours. Cells were then resuspended in a 100μl of staining buffer (PBS, 2% FBS) with a 1:200 dilution of anti-TER119-

APC (Biolegend, 116211) and incubated on ice for 30 minutes. Samples were washed three times with PBS and resuspended in 200µl staining buffer and analyzed by flow cytometry by gating on YFP- or YFP+ cells. Derivation of stable YFP-BET expressing cell lines was achieved by iterative fluorescence activated cell sorting with narrow sorting gates in order to achieve similar YFP mean fluorescent intensity between cell lines.

Whole cell extracts

Whole cell extracts were prepared using standard laboratory techniques. Briefly, 1-3e6 cells were harvested for each sample. After washing in ice-cold PBS, cells were resuspended in 50-100µl RIPA buffer (150mM NaCl, 0.5% Na Deoxycholate, 50mM Tris-HCl pH 8.0, 1% NP-40, 0.1% SDS and freshly added 500x protease inhibitor cocktail [Sigma, P8340] and 100x 100mM phenylmethylsulfonyl fluoride [PMSF]) and incubated on ice for 30 minutes. Lysed cells were sonicated in a Qsonica sonicator (model Q800R) for 3x 30 second on/off cycles at maximum amplitude in a water bath chilled to 4C. Insoluble material was pelleted by centrifugation at 15000 rpm for 5 minutes at 4C in a microcentrifuge. The supernatant was transferred to fresh microcentrifuge tubes and used immediately or snap frozen in liquid nitrogen and stored at -80C. Protein concentration was determined using the BCA protein assay (Pierce, 23225) with absorbance measured at 562nM compared to a BSA protein standard curve.

Western blotting

Western blots were performed using standard laboratory techniques. Briefly, protein samples were resuspended in 4x sample buffer (500mM Tris-HCl

pH 6.8, 8% sodium dodecyl sulfate, 20% 2-mercaptoethanol, 0.3% bromophenol blue, 30% glycerol) and denatured at 95C for 10 minutes. Equal protein amounts were loaded and separated on 4-15% or 4-20% gradient pre-cast gels (Bio-Rad, 4561084 or 4561096) and transferred to nitrocellulose membranes (Bio-Rad, 1620115). Primary antibodies include: Brd2 N-term (Cell Signaling, 5848S), Brd2 C-term (Bethyl, A302-583A), Brd3 (Active Motif, 61489), Brd4L (Bethyl, A301-985A), Brd4 N-term (Abcam, ab128874), HA (monoclonal 12CA5, purified in-house) or HA (Cell Signaling, C29F4), GFP/YFP (Abcam, ab290), PCNA (Santa Cruz, sc-9857) and β-Actin−Peroxidase (Sigma, A3854). Secondary staining reagents include goat anti-rabbit IgG-HRP (Abcam, ab6721), donkey anti-goat IgG-HRP (Santa Cruz, sc-2020) and Protein A-HRP (Thermo Fisher Scientific, 101023). Stained membranes were imaged using enhanced chemiluminescent substrate (Thermo Scientific, 34080 or 34095) and autoradiography film (Denville Scientific, E3012). For fluorescent Western blots, fluorescent secondary antibodies conjugated to IRDye-680 or IRDye-800 and the Odyssey imaging system and buffers were used (LI-COR).

Coiled coil prediction

Coiled coil prediction was performed by inputting the UniProtKB amino acid sequences for the following mouse BET proteins BRD2 (Q7JJ13-1), BRD3 (Q8K2F0-1), BRD4S (Q9ESU6-2) and BRD4L (Q9ESU6-1) into a Hidden Markov Model coiled coil prediction algorithm (https://toolkit.tuebingen.mpg.de/#/tools/marcoil)[143]. To generate the plots in Figure 3.7.A, the output was graphed and scaled to the BET protein diagrams

using ggplot in R. The sequence alignments of the regions surrounding and containing the ET and CC domains in Figure 3.7.B was generated using Jalview software[144] for the indicated UniRef entry names and fragments. Sequences were sorted by length and individual residues were labeled using Clustal X default coloring and thresholding.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed using the following antibodies: GFP/YFP (Abcam, ab290), LEO1 (Bethyl, A300-175A), CSNK2A1 (Bethyl, A300-198A). Antibody specificity was verified by comparing to isotype control IgG/sera as well as in the case of the GFP/YFP antibody to epitope-negative samples (data not shown).

For the LEO1_r1 sample, sonication-based ChIP was performed on G1E-ER4 cells as described[131]. For the YFP-BET, CSNK2A1, and LEO1_r2 samples, the ChIP protocol was adapted for enzymatic fragmentation as follows. G1E-ER4 cells (≥10 million per sample) were fixed in 1% formaldehyde in PBS at room temperature with agitation for 10 minutes, then quenched with 1M glycine for 5 minutes. Fixed cells were resuspended in 1mL cell lysis buffer (10mM Tris-HCl pH 8.0, 10mM NaCl, 0.2% NP-40/Igepal) prepared fresh with 500x protease inhibitors (Sigma, P8340) and 1mM PMSF and incubated on ice for 10 minutes. Nuclei were pelleted and digested with 10U/µl MNase in 400µl MNase buffer (NEB, M0247S) at 37C for exactly 5 minutes. MNase activity was quenched with 10µl of EGTA. Nuclei were pelleted and resuspended in 1mL nuclear lysis buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 1% SDS, prepared fresh with protease inhibitors and PMSF).

Samples were split into 500µl aliquots in 1.5mL polystyrene sonication tubes (Active Motif, 53071) and sonicated for three 10 second intervals at 4C. After sonication, samples were spun at 13000rpm in a microcentrifuge for 5 minutes at 4C to remove debris. Aliquots were combined and diluted with 4mL of IP dilution buffer (20mM Tris-HCl pH 8.0, 2mM EDTA, 150mM NaCl, 1% Triton X-100, 0.01% SDS prepared fresh with protease inhibitors and PMSF). To preclear, 50µg of protein A/G agarose beads (agarose beads slurry was prepared by mixing Protein A [Invitrogen 15918014] and Protein G [Invitrogen 15920010] agarose beads at 1:1 ratio) and 50µg of isotope-matched IgG were added. Samples were precleared for ≥2 hours. Prior to setting up immunoprecipitation ("IP") reactions, 200µl of precleared chromatin was removed as "Input." IP reactions were performed by adding 35µl of protein A/G beads pre-bound with antibody (35µl protein A/G bead slurry, 1mL PBS, 10µg antibody, incubated with rotation at 4C for ≥2 hours) and rotated overnight at 4C.

Beads were washed once with IP wash buffer 1 (20mM Tris-HCl pH 8.0, 2mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS), twice with high salt buffer (20 mM Tris-HCl pH 8.0, 2mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.01% SDS), once with IP wash buffer 2 (10 mM Tris-HCl pH 8.0, 1mM EDTA, 0.25 M LiCl, 1% NP-40/Igepal, 1% Nadeoxycholate), and twice with TE (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8). All washes were performed on ice. Following the final wash, beads were moved to room temperature and eluted twice with 100µl of elution buffer (100mM NaHCO3, 1%SDS, prepared fresh) for a final eluate volume of 200µl. The following were added to each IP and input sample: 12µl of 5M NaCl,

2µl RNaseA (10mg/ml, 10109169001 BMB) and samples were incubated at 65C for ≥1 hour. 3µl of Proteinase K (20mg/ml, 3115879 BMB) was added and samples were incubated at 65C overnight. Following overnight incubation, 10µl of 3M sodium acetate pH 5.0 was added to each sample and DNA was purified using the QIAquick PCR Purification kit (Qiagen 28106) per the manufacturer's instructions. IP samples were eluted with 60µl water, and input samples with 133.3µl water. ChIP-qPCR was performed as described above. Standard curves were constructed for each input sample and used to calculate the IP quantities for each primer set.

ChIP sequencing

Library preparation and sequencing

Samples for ChIP sequencing (ChIP-seq) were prepared as above for ChIP-qPCR. ChIP-seq libraries were processed using the TruSeq Sample Preparation Kit (Illumina, IP-202-1012/IP-202-1024) per manufacturer's instructions. Following library preparation, SPRIselect beads (Beckman Coulter, B23318) were used at 0.9X and 0.5X for left and right-side size selection, respectively, for the MNase-fragmented samples. For sample LEO1_r1, which was fragmented using sonication, 0.6X right-side size selection was used yielding library sizes in the 300-600 bp range with average sizes ~450 bp. Final library size and concentration was measured using BioAnalyzer (Agilent). Single-end sequencing (1 x 75 bp) was performed on the Illumina Nextseq 500 in high-output mode per manufacturer's instructions. Reads were converted to fastq using bclfastq2 v2.15.04 (default

parameters). Fastq reads were aligned to the mm9 genome using bowtie (bowtie --best --strata --chunkmbs 512 -S -q -l 40 -n 1 --phred33-quals -p 3 -m 1).

Computational analysis

After alignment, peaks were called and bedgraph files were generated using MACS2 (callpeak -B -p 1e-5 --SPMR --nomodel --extsize 300 --local 100000 -g mm). The bedGraphToBigWig tool from UCSC was used to convert into bigwig. Enrichment scores in peak coordinates were calculated using multiBigwigSummary in deepTools. All enrichment scores and bigwig files are library size normalized using reads per million. Merging of bigwig files for technical replicates was performed using bigwigCompare in deepTools (bigwigCompare –operation mean -bs 200). Pearson coefficients comparing correlations across samples were calculated using plotCorrelation in deepTools. Metaprofile plots across peaks were generated using computeMatrix (computeMatrix reference-point –referencePoint center -b 1000 -a 1000) and plotProfile in deepTools. Consensus peak sites were calculated from ChIPseq replicates using bedtools intersect. Differential binding analysis was performed using DiffBind. Annotation of genomic regions was performed using annotatePeaks.pl mm9 in Homer. Genome browser views were generated using IGV. Transcription factor motif analysis was performed using findMotifsGenome.pl mm9 -size 200 in Homer. A peak file containing consensus peaks generated by DiffBind was used as background to calculate motif enrichment. Scatter plots and volcano plots were generated using ggplot2 in RStudio.

Structural assessment of human BRD2/3-CC

Protein purification

The coiled coil domain of human BRD2 (715-757) and that of human BRD3 (641-688) were cloned into pGEX6p and expressed with a GST tag. BL21 (DE3) colonies transformed with respective plasmids were inoculated into LB and after overnight incubation, were used to inoculate 1-2L of culture. The culture was induced at OD600 = 0.75 with 0.5mM isopropyl β- d-1-thiogalactopyranoside (IPTG), followed by overnight expression at 25C.

The cell pellet was resuspended in 50 mM Tris pH 7.2, 500 mM NaCl, 0.1% Triton X-100, 1 mM dithiothreitol (DTT), 10µg/ml DNase I, 10 µg/mL RNase A and 1× cOmplete EDTA-free protease inhibitor (Roche, 11873580001) at ~10 mL/g of cells and then lysed via sonication. After clarification at 15,000× g for 30 min at 4C, the clear lysate was incubated with pre-equilibrated glutathione-Sepharose 4B beads (GE Healthcare, 17075601) for 1.5 hours at 4C with mixing. The beads were washed with 3× 5 CV of GST wash buffer containing 10mM Tris pH 7.2, 100 mM NaCl and 1 mM TCEP. The proteins were eluted with wash buffer containing 50 mM reduced glutathione. The GST tag was cleaved using HRV-3C protease by incubating with the eluted protein overnight. On the next day, the cleavage protein sample was concentrated and then injected onto a pre-equilibrated (in the GST wash buffer) Superdex™ 75 16/600 column, and the gel-filtration chromatography was run at 1 mL/min for 1.25× CV with 1-mL fractions collected. Fractions containing desired bands were combined and concentrated to 700 µM for SEC-MALLS analysis.

Size exclusion chromatography

Size exclusion chromatography with inline multi-angle light scattering (SEC-MALLS) analysis runs were performed on either a Superdex™ 75 10/300 GL column or a Superdex™ peptide 10/30 GL column (GE Healthcare) with an in-line MiniDawn MALLS detector with a laser source at 690 nm (Wyatt Technology, Santa Barbara, CA, USA) and Wyatt refractometer. Proteins were eluted in 10 mM Tris pH 7.2, 100 mM NaCl, 1mM TCEP using a flow rate of 0.5 mL/min. The weight-average molecular weight was calculated using the intensity of scattered light at 90° in combination with the change in refractive index. Protein concentration at the detector was determined by the change in refractive index. Allexperiments were performed at room temperature (25C). Data collection and SEC-MALS analysis were performed with ASTRA 6.1 software (Wyatt Technology).

Circular dichroism spectropolarimetry

Circular dichroism spectropolarimetry spectra were recorded on a Jasco J-720 spectropolarimeter using a 1-mm quartz cuvette. In each case, spectra comprised the sum of three successive spectra with a step size of 0.5 nm, a 1-s response time and a 1-nm bandwidth. Data were acquired at 4C. Secondary structure content was estimated using Raussens' method (http://perry.freeshell.org/raussens.html).

GST pulldown and proteomic analysis

Protein purification

The sequences from mouse BRD2, BRD3 and BRDT (BRD2-mB, BRD2-ET, BRD2-CC, BRD2-CCp, BRD2-CCd, BRD2-ETCC, BRD3-ET, BRD3-ETCC, BRDT-ET, and BRDT-ETCC [Table 2.2]) were cloned into the pGEX-2T GST

fusion vector. The GST fusion proteins were expressed and purified from bacteria as above with the following changes: after BL21 transformation, a 5mL LB/Amp culture grown at 37C overnight was used to inoculate 200mL LB/Amp, which was then grown at 37C for 2.5 hours. Protein expression was induced with 0.25mM IPTG and the culture incubated at 30C for an additional 2.5 hours. Cell pellets were resuspended in 5mL BC500 (20mM Tris-HCl pH 8.0, 0.5mM EDTA, 50mM KCl, 20% glycerol, 1% NP-40, 1mM DTT, 500x protease inhibitor cocktail [Sigma, P8340]). Prior to sonication, 50µl of 100mg/mL lysozyme and 0.5mL of 20% Triton X-100 was added. Bacterial suspensions were sonicated (Misonix, Sonicator 3000) at 50% output with microtip limit for 1 second for three sets of 10 cycles, keeping samples on ice. The bacterial lysates were cleared with 10,000rpm centrifugation for 10 minutes at 4C. The cleared supernatant was harvested and 250µl of glutathione agarose beads (Pierce, 16100) were added with gentle inversion overnight at 4C. After overnight incubation, the beads were centrifuged at 1500rpm at 4C for 3 minutes. The beads were washed 4x in 5mL BC500 and 1x in a wash buffer (1x PBS, 20% glycerol, 1% NP-40, 1mM DTT, 500x protease inhibitor cocktail [Sigma, P8340]). GST proteins were eluted from the beads using 3x washes in 0.833 mL reduced glutathione solution (50mM Tris-HCl pH 8.8, 100mM NaCl, 20mM reduced glutathione [Sigma, G4251], final pH adjusted to 8.0) pooling each eluate yielding approximately 2.5mL of soluble protein. Protein samples were concentrated using Amicon Ultra-4 conicals (Millipore-Sigma) centrifuged at 4100rpm for 10-40 minutes per manufacturer's instructions. The spin was repeated twice after refilling to maximum volume with BC100 (20mM Tris-HCl

pH 8.0, 100mM KCl, 0.5mM EDTA, 20% glycerol, 1mM, DTT, 500x protease inhibitor cocktail [Sigma, P8340]). Protein concentration was estimated using BSA standards on an SDS-PAGE gel and aliquots were stored at -80 °C.

Nuclear extract preparation

Nuclear extracts were prepared from murine erythroleukemia (MEL) cells. Briefly, ~100e6 cells increments were washed in PBS and resuspended in 10x cell pellet volume of Buffer A (10mM HEPES-KOH pH 7.9, 1.5mM MgCl2, 10mM KCl, 1mM DTT, 100mM PMSF, 500x protease inhibitor cocktail [Sigma, P8340]) and incubated on ice for 30 minutes with intermittent vortexing to lyse the cell membrane. Nuclear pellets were resuspended in 3x pellet volume of Buffer C (20mM HEPES-KOH pH 7.9, 25% glycerol, 420mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 1mM DTT, 100mM PMSF, 500x protease inhibitor cocktail [Sigma, P8340]) and incubated on ice for 30 minutes with intermittent vortexing. Nuclear extracts were cleared by spinning in tabletop microcentrifuge at 13000rpm for 5 minutes at 4C. Concentrations for cleared extracts were determined using the A280 absorbance measurement on a spectrophotometer (Thermo Scientific, Nanodrop 2000). Samples were stored at -80C.

GST pulldown assay

GST pulldown assays were performed by diluting 500µg of nuclear extract in dilution buffer (50mM Tris-HCl pH 7.5, 0.5% NP-40) to a final NaCl concentration of 150mM and bringing to a final volume of 1mL using binding/wash buffer (150mM NaCl, 50mM Tris-HCl pH 7.5, 0.5% NP-40). 2.5µg of respective GST fusion proteins and 20µl of glutathione agarose beads (Pierce, 16100) were added and

samples were incubated with rotation 4 hours to overnight at 4C. Beads were washed 5x in 1mL binding/wash buffer and eluted in 50µl 4x sample buffer. Eluted samples were run on SDS-PAGE gels and either submitted for mass spectrometry or analyzed by Western blot (see above).

Mass spectrometry

Liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) analysis was performed by the Proteomics and Metabolomics Facility at the Wistar Institute using a Q Exactive Plus mass spectrometer (Thermo Fisher) coupled with a Nano-ACQUITY UPLC system (Waters). Samples were digested in-gel with trypsin and injected onto a UPLC Symmetry trap column (180 µm i.d. × 2 cm packed with 5 µm C18 resin; Waters). Tryptic peptides were separated by reversed-phase HPLC on a BEH C18 nanocapillary analytical column (75 µm i.d. × 25 cm, 1.7 µm particle size; Waters) using a 95 min gradient formed by solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile). A 30 min blank gradient was run between sample injections to minimize carryover. Eluted peptides were analyzed by the mass spectrometer set to repetitively scan *m*/*z* from 400 to 2,000 in positive-ion mode. The full MS scan was collected at 70,000 resolution followed by data-dependent MS/MS scans at 17,500 resolution on the 20 most abundant ions exceeding a minimum threshold of 20,000. Peptide match was set as preferred, exclude isotopes option and charge-state screening were enabled to reject singly and unassigned charged ions.

Proteomic analysis

Peptide sequences were identified using MaxQuant 1.5.2.8. MS/MS spectra were searched against a UniProt mouse protein database using full tryptic specificity with up to two missed cleavages, static carboxamidomethylation of Cys, and variable oxidation of Met and protein N-terminal acetylation. Protein quantification was performed using razor +unique peptides. Razor peptides are shared (non-unique) peptides assigned to the protein group with the most other peptides (Occam's razor principle). False discovery rates for protein and peptide identifications were set at 1%. Identified proteins required a minimum of 2 razor or unique peptides in at least one of the samples. The list of proteins was further filtered by dropping proteins not identified in 5/7 observed samples per replicate and by dropping proteins observed in only one of three replicates per sample. Intensity values were normalized to GST detected for each sample. Values for proteins absent in one replicate were imputed by averaging the intensity values of the flanking proteins in the other two replicates. If the flanking proteins are absent in one of the two remaining replicates, the protein was discarded, and if one of the flanking proteins was absent in both replicates, an average of the two remaining flanking proteins (one from each replicate) was used. Finally, ComBat[145] was applied to reduce batch effect between replicates. Comparison between samples was performed by calculating the mean of the log2 of the ratio between two samples from each replicate and the log10 of the Q-value. STRING protein interaction network analysis (https://string-db.org/) was performed on proteins more strongly enriched in the ETCC sample than the ET sample (log2ratio > 1.2,

logQvalue > 4) using the highest confidence setting (0.900) labeled with experimentally validated interactions and interactions from curated databases.

Data availability

The FASTQ files for the RNA sequencing data have been deposited in the NCBI BioProject database with BioProject accession number PRJNA560407. ChIP sequencing data will be made publicly available through the Gene Expression Omnibus (GEO) pending publication.

CHAPTER 3: DOMAIN MAPPING REVEALS A COILED COIL THAT FUNCTIONALLY DISTINGUISHES BET PROTEINS

Chapter summary

Research described in this chapter forms the basis of a pending manuscript currently in the review process title “Comparative structure-function analysis of bromodomain and extraterminal motif (BET) proteins using a gene-complementation system”. Many people contributed to this work. The initial cell growth and hemoglobinization experiments were performed by Sarah Hsu and Kristen Jahn while they were in the Blobel lab. Sarah Hsu also conducted the RNA sequencing experiment in collaboration with Tapan Ganguly and Erik Toorens from the Perelman School of Medicine Penn Genomics Analysis Core. Computational analysis of the RNA sequencing data was performed by Zhe (Jim) Zhang from the Children’s Hospital of Philadelphia (CHOP) Department of Biomedical and Health Informatics. Analysis of proteomic data was carried out in collaboration with Perry Evans of the CHOP Department of Biomedical and Health Informatics. Other members of the Blobel lab contributed significantly to this work. Hongxin Wang provided a substantial amount of technical support as did Nicole Hamagami and Jennifer Yano.

Introduction

Bromodomain and extraterminal motif (BET) proteins mechanistically link chromatin acetylation to gene transcription[15,20]. The BET family consists of BRD2, BRD3 and BRD4, which are widely expressed across tissue types, and the testis-specific BRDT[20]. BET proteins share a general domain structure consisting of

tandem N-terminal bromodomains and a C-terminal ET domain conserved across species[18]. The bromodomains bind to acetylated lysine residues on histones and transcription factors[15]. The ET domain is a protein-protein interaction motif that recruits additional co-activators[31,64,66,68,71,72]. BET proteins thereby function as molecular scaffolds for transcriptional regulators. Small molecules which target the bromodomains partially displace BET proteins along with associated co-factors from chromatin[95,96]. BET inhibitors are in clinical trials for the treatment of several cancers and may have additional use in inflammatory, metabolic and cardiac disorders[20,45]. Because the bromodomains are similar between BET proteins, molecules targeting them do not strongly discriminate between individual BET proteins[95,96,117]. The lack of specificity might underlie some of the unwanted side effects of BET inhibitors and constrain their use. Moreover, clinical exploitation of BET inhibition requires a better understanding of selective BET protein functions.

Structurally, the long-isoform of BRD4 (BRD4L) is the most distinct BET protein in that it possesses a unique extended C-terminal tail that recruits factors required for the transition of RNA polymerase 2 from transcriptional initiation into productive elongation[76,78,146]. The short isoform of BRD4 (BRD4S) lacks this domain and is thus structurally more similar to BRD2 and BRD3, but its expression and function in most cell types remains poorly characterized[70,82]. Depletion of BRD2 or BRD3 results in transcriptional and phenotypic outcomes that in some settings are different from BRD4 depletion[104,124], suggesting non-overlapping functions among them. In fact, BRD2 deficient mice exhibit a phenotype distinct from that of BRD4 deficient mice[21–23,119]. BRD3-null mice have not been reported.

It has been proposed that BET-specific functions are conveyed in part via selectivity of the bromodomains for acetylated substrates[44,47,48,51–53]. Specificity might also be mediated by additional domains that influence chromatin occupancy[35,36] or by differences in protein interactions[30,31]. However, the ET protein interaction domain is highly conserved across BET proteins, making its contribution to BET-selective interactions unclear[64,66,68,71,72]. Thus, whereas all BET proteins are enriched at acetylated chromatin, the mechanisms underlying BET-specific functions remain unresolved.

GATA1-mediated erythroid cell differentiation has provided a useful model for the dissection of BET protein biology[132]. GATA1 is a master erythroid transcription factor that drives the terminal maturation of erythroid precursor cells[136]. BRD3 and to a lesser extent BRD4 bind to GATA1 in an acetylation dependent manner and co-localize with GATA1 on chromatin[113,139]. Point mutations in the BET-binding motif of GATA1, or pan-BET inhibition, impair GATA1 occupancy on chromatin[113,139,140]. GATA1 transcriptional activity has been extensively studied in G1E-ER4 cells, an erythroid progenitor line that stably expresses a conditional form of GATA1 fused to the ligand binding domain of the estrogen receptor (GATA1-ER)[147]. Using this system, GATA1-ER target genes can be activated upon treatment with estradiol, which faithfully recapitulates terminal erythroid maturation[147]. BRD2 and BRD4 are both necessary for GATA1-ER-mediated gene expression[113]. In contrast, BRD3 appears to be dispensable, but BRD3 loss exacerbates defects associated with BRD2 deficiency, suggesting overlapping functions of these two proteins[113]. Accordingly, overexpression of

BRD3 can partially compensate for BRD2 loss[113]. BRD3 is expressed approximately four-fold less than BRD2 based on RNA-sequencing in G1E-ER4 cells, potentially explaining its lack of compensatory activity when not overexpressed[148]. It is unclear whether BRD2 and BRD3 possess structural features that could facilitate functions that are distinct from BRD4.

To better understand functional differences between BRD2, BRD3 and BRD4, we took advantage of *Brd2*-null G1E-ER4 erythroblasts. These cells are viable but fail to undergo GATA1-mediated terminal maturation. We asked whether reconstitution of these cells with exogenous BET proteins rescues their growth rate, maturation and gene expression. We found that the BRD4 isoforms are functionally different from BRD2 and BRD3. Moreover, the bromodomains contribute surprisingly little in terms of BET specific functions. We define a small region adjacent to the ET domain that contains a predicted coiled coil structure in BRD2 and BRD3 but not BRD4 that we show to contribute to BRD2 selective activity.

BRD2 is essential for GATA1 dependent erythroid differentiation and gene expression

A mechanistic explanation for specific functions of individual BET proteins remains unresolved. We therefore took advantage of a cellular system in which BET proteins can be directly compared. We generated four *Brd2*-null G1E-ER4 cell lines (KO clones, KO1 - KO4) using CRISPR-Cas9 (see "Materials and Methods"). The absence of BRD2 was confirmed by Western blot with antibodies against BRD2 N- and C-terminal epitopes (Figure 3.1.A). We measured BRD3 and BRD4L protein levels and found them to be unchanged in the BRD2 KO cells (Figure 3.1.A), indicating that BRD2 is not necessary

for BRD3 and BRD4L expression, and that any defects in the BRD2 KO cells cannot be attributed to their loss. We were unable to detect endogenous BRD4S protein in whole cell lysates using commercially available antibodies (Figure 3.1.B) but reads mapping to the BRD4S-specific exon were detected by RNA sequencing. Exon-exon junction analysis suggested that BRD4S is the minor isoform with approximately ~33% of BRD4 transcripts corresponded to it in parental G1E-ER4 cells, ~24% in BRD2 KO1 cells and ~19% in KO2 cells, indicating that BRD2 does not influence which isoform predominates. Thus, changes in BRD4 mRNA alternative splicing likely do not account for any phenotypic differences observed in BRD2 KO cells. To test whether GATA1-driven erythroid maturation depends on BRD2, we exposed BRD2 KO clonal lines to estradiol for 24 hours[147], a time point sufficient for the maximal accumulation of GATA1-ER target genes. Consistent with our previous studies[113], all BRD2 KO clones failed to induce the expression of major erythroid GATA1 target genes, including *Hbb-b1*, *Alas2*, *Slc4a1 and Spta1* (Figure 3.1.C). Importantly, BRD2 depletion did not appear to impact RNA levels of the *Gata1-ER* transgene or housekeeping genes *Actb* and *Gapdh*. Consistent with the Western blots, *Brd3* and *Brd4* RNA levels were unchanged in BRD2 KO cells.

Identification of a BRD2-specific transcriptional signature

To measure the global transcriptional changes associated with BRD2 loss, we performed RNA-sequencing on parental G1E-ER4 and two BRD2 KO clones (KO1 and KO2) after 24 hours of GATA1-ER induction. Pearson correlation coefficients demonstrated strong reproducibility between the three replicates for each cell line and higher similarity of KO1 and KO2 (Figure 3.2.A). The overall high correlations between each of the samples indicate that a majority of genes are similarly expressed.

Dimensionality reduction using principle component analysis showed separation of the two KO clones from the parental G1E-ER4 cells on PC1 (37.3% of variance), revealing the transcriptional impact of BRD2 depletion. Separation of the three cell lines on PC2 (18.9% of variance) suggested clonal variation between the cell lines (Figure 3.2.B). We therefore classified differentially expressed genes as those that exhibit a ≥ 2-fold change (false discovery rate [FDR] < 0.01) in both KO clones compared to parental G1E-ER4 cells and that were unchanged between the two KO clones (Figure 3.2.C). This identified 158 downregulated genes (Table 3.1) and 261 upregulated genes (Table 3.2). A less stringent cutoff of ≥ 1.5-fold change (FDR < 0.05) resulted in 433 and 626 genes that were downregulated and upregulated, respectively (Figure 3.2.C). Expression of several highly expressed maturation associated genes (*Hbb-b1*, *Hbb-b2*, *Hba-a1*, *Hba-a2*, *Alas2*, *Slc4a1* and *Spta1*) was decreased in the BRD2 KO cells in comparison to the parental cells, whereas many genes whose expression is high in immature cells and downregulated during erythroid maturation (*Gata2*, *Myc*, *Myb*, *Kit*, *Il2rg*, *Lyl1* and *Vim*) were elevated in at least one of the BRD2 KO cell lines (Figure 3.2.D). This indicates a general failure of BRD2 KO cells to fully mature upon GATA1-ER activation.

To determine what biological pathways could be affected by the observed transcriptional changes, we queried the differentially expressed genes for enrichment of hallmark gene sets in the Molecular Signatures Database. The gene sets most associated with the downregulated genes were HEME_METABOLISM, MTORC1_SIGNALING and CHOLESTEROL_HOMEOSTASIS (Figure 3.2.E, right). These pathways are consistent with known biological processes important for erythroid maturation[149–151]. The gene sets most associated with the upregulated genes include IL2_STAT5_SIGNALING, TNFA_SIGNALING_VIA_NFKB and P53_PATHWAY. The relevance of these pathways

is not immediately clear but could reflect a failure to downregulate genes important for cell proliferation. An analysis using the less stringent list of differentially expressed genes (fold change ≥ 1.5, FDR <0.05) yielded similar gene set enrichment results (not shown).

The pharmacologic BET inhibitor JQ1 targets all BET proteins with comparable efficacy[95]. We assessed to what extent the cellular response to JQ1 can be attributed to BRD2 inhibition by comparing the transcriptome analyses of BRD2 KO cells to our prior data sets in which these cells were exposed to JQ1. Since the prior experiments employed a microarray platform[113], we first intersected genes detected by both methods, resulting in 9132 genes (see "Materials and Methods"). Comparison of the gene expression changes in the two datasets revealed a positive correlation between genes impacted by BRD2 loss and JQ1 exposure (Pearson correlation coefficient of R=0.47 ($p<2.2e^{-16}$)) (Figure 3.2.F). Of the 248 genes upregulated in BRD2 KO cells, 104 were also upregulated (FC>2, FDR<0.01) upon JQ1 treatment ($p<2.2e^{-16}$), and of the 137 downregulated genes, 54 were also downregulated upon JQ1 treatment ($p<2.2e^{-16}$) (Figure 3.2.G). A limitation to this comparison is that alterations in gene expression upon chronic BRD2 loss are a composite of direct and indirect effects. In contrast, JQ1 treatment enables detection of an immediate response to BET inhibition. Moreover, JQ1 does not fully displace BET proteins from chromatin[95,117], which might limit effect size. In spite of these limitations, these results suggest that impaired BRD2 function accounts for at least part of the effects of JQ1 treatment. In sum, BRD2 is critical for the regulation of a wide array of erythroid genes involved in multiple aspects of GATA1-mediated differentiation, and loss of BRD2 activity accounts for at least part of the JQ1 effects.

BET specific functions during erythroid cell growth and differentiation

While BET proteins are structurally similar, few studies have directly examined whether conserved protein domains contribute to BET selective functions or whether they simply reflect functional overlap (Figure 1.4, Table 1.1)[25,35,79,85,152]. We previously reported that BRD2 and BRD3 can partially compensate for each other in G1E-ER4 cells[113]. We next tested whether a similar functional overlap exists between BRD2 and either BRD4S or BRD4L in these cells. Although BRD4S was not detected in G1E-ER4 cells, we included it in this functional comparison due to its structural resemblance to BRD2 and BRD3. We used a murine retroviral expression vector to transduce HA-tagged forms of BRD2, BRD3, BRD4S or BRD4L cDNA into parental G1E-ER4 cells or BRD2 KO1 cells. We used three assays to study the relative activities of these proteins: 1) a reddening of the cells indicating activation of the hemoglobin synthesis pathways, 2) expression of erythroid maturation associated genes upon GATA1-ER activation, and 3) cell growth. To visualize hemoglobinization, we induced GATA1-ER activity for 48 hours at which time the cells normally have turned red. As expected, BRD2 was required for hemoglobin production (Figure 3.3.A). Expression of BRD2 restored the ability of BRD2 KO1 cells to produce hemoglobin. BRD3 was also able to rescue hemoglobinization, but both BRD4S and BRD4L displayed much lower activity in this assay. We next measured the mRNA levels of *Hbb-b1*, which was one the most differentially regulated genes in the BRD2 KO cells based on the previous analysis. Activation of GATA1-ER failed to induce expression of *Hbb-b1* in the BRD2 KO1 cells. However, expression of BRD2 and BRD3 restored *Hbb-b1* expression in BRD2 KO1 cells to similar levels while BRD4S and BRD4L showed little activity (Figure 3.3.B). Note that overexpression of BET proteins in parental cells did not overtly perturb *Hbb-b1* expression or hemoglobin production, suggesting that excess levels of a given BET protein is compatible with cell maturation.

To further compare the activities of BET proteins, we assessed the impact of BET expression on cell growth. The retroviral vector encoding the BET cDNAs contained an IRES-GFP module to mark cells with stably integrated constructs. We designed a competitive growth assay which monitors the fraction of GFP+ cells in a mixture of infected and uninfected cells over time using flow cytometry. Restoration of BRD2 expression in BRD2 KO1 cells increased the fraction of GFP+ cells by ~40% over the time course (Figure 3.3.C). BRD3 expression augmented growth almost to the same extent as BRD2. In contrast, BRD4S expression failed to rescue BRD2 KO1 cell growth and even lowered cell numbers. We also measured growth rates in cells expressing BRD4L and found BRD4L to be incapable of restoring BRD2 KO1 cell growth (Figure 3.3.D). In parental cells, BRD2 overexpression had no impact on cell growth, whereas BRD3 resulted in a ~20% decrease, BRD4S in a ~40% decrease and BRD4L in a 10% decrease (Figure 3.3.C,D), suggesting that excess levels of these molecules can impair proliferation or viability. In sum, both differentiation and growth assays point at overlapping functions between BRD2 and BRD3, while BRD4S, which is overall structurally similar, and BRD4L exert distinct functions in these cells.

The abundance of BRD2 and BRD4S is strongly influenced by their N-terminal halves

Even though the BET constructs were expressed by the same vector, the resulting protein amounts were markedly variable as determined by Western blot with anti-HA antibodies (Figure 3.3.E). Specifically, BRD3 and BRD4S were readily detected, whereas BRD2 was not detected unless a BRD2-specific antibody was used. Differences in exogenous BET expression potentially confound interpretation of the results. However, comparisons between the effects of BRD3 and BRD4S might still be instructive since these proteins are expressed at similar levels and their effect on BRD2 comparable. For

example, BRD3 and BRD4S exert opposite effects on BRD2 KO1 cell growth, with BRD3 restoring their expansion and BRD4S lowering cell numbers (Figure 3.3.C). Similarly, whereas BRD3 rescued *Hbb-b1* expression to levels comparable to that achieved by BRD2, BRD4S failed to do so (Figure 3.3.B).

Variable production of BET proteins could be due to differences in mRNA stability, translation and/or protein turnover. To address this, we measured mRNA levels of the BRD2, BRD3, BRD4S and BRD4L transgenes in the above cells using primers common to all constructs (specific to the GFP coding portion of the mRNAs). We found that mRNA levels varied substantially less than protein levels. In fact, *Brd2* transcripts were somewhat higher than the others (Figure 3.3.F), indicating that the decreased BRD2 protein was due to a post-transcriptional mechanism.

To identify which features may be dictating the contrasting nature of BRD2 and BRD4S in terms of function and stability, we began to dissect their differences by designing chimeric BET constructs consisting of the N-terminal and C-terminal halves of BRD2 and BRD4S fused immediately downstream of the second bromodomain. These chimeric BET proteins were termed BRD2-4S and BRD4-2, respectively (Figure 3.4.A). This strategy preserved their overall modular domain structure, including the N- and C-terminal nuclear localization signals[153,154]. These constructs were introduced into the same retroviral vector as described before. We infected parental and two BRD2 KO cell lines expressing the chimeric and intact BET proteins and isolated stably expressing cell lines. We measured mRNA levels produced by each construct and found them to be similar (Figure 3.4.B). Strikingly, the protein levels of BRD2 and the chimeric BRD2-4S were dramatically lower than those of BRD4S and chimeric BRD4-2 (Figure 3.4.C). To quantify these differences in protein expression, we repeated the infection of BRD2

KO1 cells with retroviruses encoding the intact BRD2, BRD3, BRD4S and BRD4L HA-tagged proteins, as well as the chimeric BRD2-4S and BRD4-2 constructs and performed a Western blot probing the HA epitope using fluorescently labeled antibodies. This strategy allowed the visualization and quantification of all six BET proteins (Figure 3.4.D). Consistent with our prior observations, BRD2 and BRD2-4S were expressed at equally low levels in comparison to BRD3, BRD4S and BRD4-2. BRD4L was expressed at an intermediate level. This indicates that steady state protein levels are determined by the N-terminal halves of the BET proteins.

Selective functions of BET proteins are determined by their C-terminal halves

To determine whether there are BET-selective functions that are conveyed by domains that do not affect protein levels, we tested the ability of BRD2, BRD4S and the chimeric BRD2-4S and BRD4-2 proteins to restore BRD2 KO cell growth and differentiation-associated gene expression. BRD2-4S, which was expressed at similar levels as BRD2, was impaired in its ability to restore cell growth compared to BRD2 (Figure 3.4.E), whereas BRD4-2, which was expressed similarly to BRD4S, had a positive effect on BRD2 KO cell growth (Figure 3.4.F). Given that BRD2 but not BRD4S rescued the expression of *Hbb-b1* in BRD2 KO cells (Figure 3.3.B), we asked whether the C-terminal halves of BRD2 and BRD4S were responsible for this differential activity. BRD4-2 rescued *Hbb-b1* expression comparably to wild type BRD2 whereas BRD2-4S had no measurable effect (Figure 3.4.G). The same trends held true for the expression of the other major globin chain gene *Hba-a1*, though compared to *Hbb-b1*, *Hba-a1* expression was less affected by BRD2 depletion (Figure 3.4.H). Together, these observations suggest that the C-terminal half of BRD2 encodes its specific activity, which can be grafted onto the N-terminal half of BRD4S to substantially convert it into a BRD2-like molecule. This further

suggests that the specific functions of BET proteins are not critically determined by differences in the bromodomains, but rather by differences in the C-terminal halves of the proteins.

So far, comparisons among proteins were largely limited to those that were expressed at comparable levels, but the drastic differences in protein abundance conferred by the N-terminal halves of BRD2, BRD3 and BRD4S confounded direct functional comparison across all of the constructs. To overcome these limitations, all BET constructs were fused to YFP and stably introduced into BRD2 KO1 cells via retroviral transduction. Generation of cell lines with comparable expression of each construct required careful titering of viruses and multiple rounds of fluorescence-activated cell sorting using stringent gating on the YFP signal (Figure 3.5.A). In that manner we eventually succeeded in isolating subpopulations of cells in which all proteins were produced at nearly equal levels that in turn were similar to the amount of endogenous BRD2 in parental G1E-ER4 cells (Figure 3.5.B). YFP-BET transgene mRNAs were measured with qPCR primers that target the *YFP* sequence (*YFP*) and the 5' and 3' end of *Brd2* (Figure 3.5.C). These measurements indicate that the RNA levels corresponding to the YFP-BRD2 and YFP-BRD2-4S proteins were slightly higher than the RNA levels for the YFP-BRD3, YFP-BRD4S and YFP-BRD4-2 proteins. This is consistent with the lower protein yields for BRD2 and BRD2-4S, requiring selection of cells with higher RNA expression. These relative mRNA differences, however, were not as drastic as the original protein differences. The YFP sequence may stabilize BRD2 mRNA or protein to some extent resulting in a better correlation between its RNA and protein levels. Of note, levels of *Gata1-ER* and *Gapdh* mRNA were similar in the cell lines (Figure 3.5.C).

We next assessed whether the reconstituted BRD2 KO1 cells could activate eight BRD2-dependent genes during erythroid maturation. As expected, expression of *Hbb-b1, Alas2*, *Slc4a1*, *Spta1, Csf2rb, Csf2rb2, Kcnn4* and *Rhag* were all reduced in the un-reconstituted BRD2 KO1 cells compared to the parental G1E-ER4 cells (Figure 3.5.D). Expression was rescued by YFP-BRD2 and to a lesser extent by YFP-BRD3, indicating that the increased bulk added by the YFP fusion onto the N-terminus of BET proteins did not impact their overall function. YFP-BRD4S had a marginal impact on the expression of these BRD2-dependent genes and was clearly less potent than YFP-BRD3. We next assessed the functions of the chimeric YFP-BRD2-4S and YFP-BRD4-2 BET proteins. YFP-BRD2-4S was unable to restore expression of 7 of the 8 genes. In contrast, YFP-BRD4-2 was able to restore expression of all 8 genes to levels comparable to YFP-BRD2 and parental G1E-ER4 cells.

To test for the global rescue of GATA1-induced erythroid maturation by the various BET proteins, we examined the ability of the YFP-BRD2, YFP-BRD4S and the two chimeric YFP-BRD2-4S and YFP-BRD4-2 proteins to rescue hemoglobinization in BRD2 KO1 and KO2 cells after sorting for equal YFP-BET levels, which was confirmed by post-sorting flow cytometry revealing comparable YFP intensity for all of the YFP-BET proteins (Figure 3.6.A). As previously observed, parental G1E-ER4 cells fully hemoglobinized whereas the BRD2 KO1 and KO2 cells failed to turn red (Figure 3.6.B). YFP-BRD2 expression in both BRD2 KO cell lines rescued hemoglobin production to levels comparable to the parental G1E-ER4 cells, whereas YFP-BRD4S expressing BRD2 KO cells failed to turn red. Chimeric YFP-BRD2-4S failed to rescue BRD2 KO cells whereas YFP-BRD4-2 restored hemoglobinization comparably to, if not better, than YFP-BRD2. To further confirm that these phenotypic changes reflect underlying gene expression changes

associated with rescued GATA1-ER activity, we measured mRNA levels of candidate BRD2-dependent GATA1-activated genes in these cells after 24 hours of GATA1-ER activation. These data indicate that the C-terminal but not the N-terminal half of BRD2 conveys BRD2's functional specificity in BRD2-dependent cell growth, differentiation and gene expression.

BRD2 and BRD3 contain a putative coiled coil domain that contributes to BRD2 activity

The previous experiments reveal that BRD2-like activity shared by BRD2 and BRD3 maps to the C-terminal half of BRD2. We therefore examined the amino acid sequences of the C-terminal halves of both BRD2 and BRD3 for shared features that may distinguish them from BRD4S. Coil prediction of the mouse BET proteins using MARCOIL[143] reveals multiple putative coiled coils (Figure 3.7.A). We identified a region immediately downstream of the ET domain that is predicted to harbor a coiled coil (that we refer to as the CC domain) in BRD2 and BRD3 but not in BRD4. The CC domain is conserved across the BRD2, BRD3 and BRDT paralogues in human, mouse and zebrafish (Figure 3.7.B). It is also conserved in the *Drosophila* BET paralog FSH as well as one of the yeast paralogs, BDF1 but not BDF2. We therefore asked whether this region could account for some of the observed differences between BRD2/BRD3 and BRD4S.

To expedite analyses of several constructs testing the function of this region without having to generate multiple independent cell lines for each, we developed a flow cytometry-based version of the erythroid maturation gene complementation assay in G1E-ER4 cells. This approach quantifies the fluorescent tag on each construct as a proxy for protein abundance and the erythroid-specific surface antigen TER119 as a measure of GATA1-mediated differentiation (Figure 3.8.A). As expected, TER119 was only expressed on G1E-ER4 cells upon GATA1-induced maturation (Figure 3.8.B)[155]. We

validated the use of the TER119-based erythroid maturation assay by demonstrating that YFP-BRD2, but not a YFP empty vector, restored GATA1-mediated TER119 expression in BRD2 KO cells to levels observed in the parental G1E-ER4 cells (Figure 3.8.B, 3.8.C). We also compared erythroid gene expression levels to TER119 surface staining after 24 hours of GATA1-ER activation in parental G1E-ER4 cells or BRD2 KO1 cells expressing YFP-fused BRD2, BRD3, BRD4S, or chimeric BRD2-4S or BRD4-2 (Figure 3.8.D). There was a strong correlation between the amount of TER119 expression on the surface of these cells and the mRNA levels of the erythroid genes *Alas2* (R=0.81, p=0.026), *Hbb-b1* (R=0.84, p=0.017*)*, *Slc4a1* (R=0.82, p=0.024) and *Spta1* (R=0.88, p=0.009). For example, whereas the parental G1E-ER4 cells exhibited high levels of mRNA and TER119 staining, the BRD2 KO1 cells without YFP-BET contained low mRNA levels of these genes and exhibited minimal TER119 staining. However, the YFP-BRD2-rescued BRD2 KO1 cells expressed mRNA and TER119 levels comparable to the parental G1E-ER4 cells. YFP-BRD3 and YFP-BRD4-2 likewise rescued the BRD2 KO1 cells, albeit to varying degrees, whereas YFP-BRD4S and YFP-BRD2-4S failed to rescue mRNA and TER119 expression. These data indicate that the gene encoding the TER119 antigen is BRD2-dependent and is representative of erythroid maturation.

The ET region is a protein-protein interaction domain that recruits transcriptional co-activators to chromatin[64]. Given the proximity between the CC and ET domains, it is possible that the CC contributes to the ET-mediated interactions or functions independently. We therefore deleted both the ET (BRD2dET) and CC (BRD2dCC) regions individually to determine if either region contributes to BRD2 activity in G1E-ER4 cells using YFP-BET fusions (Figure 3.9.A, middle). YFP-BRD2dET and YFP-BRD2dCC were expressed at levels similar to full length YFP-BRD2 based on YFP fluorescence (Figure

3.9.A, left). Deletion of the ET domain completely abrogated TER119 expression in comparison to full length YFP-BRD2 (Figure 3.9.A, right). Deletion of the CC domain also reduced TER119 expression, albeit to a lesser extent. To measure the effect of the ET and CC deletions on BRD2 protein expression and BRD2-mediated gene expression and cell growth, we expressed HA-tagged versions of these proteins in BRD2 KO1 cell lines. RNA levels of the constructs were similar based on qPCR primers targeting the 5' end of *Brd2* and the *GFP* sequence (Figure 3.9.B). Levels of *Gata1-ER, Brd3,* and *Brd4,* as well as the control genes *Actb* and *Gapdh* were similar across the cell lines. The protein levels of the constructs were likewise similar, indicating that the ET and CC domains do not significantly contribute to BRD2 stability (Figure 3.9.B, inset). As seen previously, restoration of BRD2 rescued the expression of multiple BRD2-dependent genes (*Hbb-b1*, *Alas2*, *Slc4a1*, *Spta1, Csf2rb2, Kcnn4, Rhag*) (Figure 3.9.C). Deletion of the ET domain completely abrogated the ability of BRD2 to rescue these genes. Deletion of the CC domain had a more partial, but uniform, effect on the expression of these genes. These data indicate that whereas the ET domain is absolutely critical for proper BRD2 functioning during erythroid maturation, the CC region contributes more partially to BRD2 function.

In addition to restoring erythroid gene expression, BRD2 also restores the cell growth rate of BRD2 KO cells. We therefore tested whether the ET and CC regions contributed to BRD2-mediated cell growth. As performed previously to measure the effect of BET protein expression on cell growth, we infected parental G1E-ER4 as well as BRD2 KO1 and KO2 cells with viruses encoding HA-tagged BRD2, BRD2dET and BRD2dCC and tracked the effect of BET expression on cell growth by measuring the fraction of the GFP+ cells over time. Reconstitution of BRD2 in BRD2 KO1 and KO2 cells led to a ~45% increase over the time course (Figure 3.9.D). Deletion of the ET domain drastically

decreased BRD2 function in this assay. Interestingly, though the CC domain contributed only partially to erythroid gene expression as seen above, its deletion had a significant impact on BRD2 function in cell growth, indicating that the erythroid genes assayed may not completely reflect the overall gene expression changes associated with loss of the BRD2 CC region. In summary, whereas the ET domain is essential for BRD2 function, the CC region may play a more variable role in the expression BRD2-dependent genes.

The combined ETCC region functionally distinguishes BRD2 and BRD4S

Like the bromodomains, the ET region is highly conserved between BET proteins. The adjacent CC region, however, is present in BRD2 and BRD3 but not in BRD4. We therefore asked whether either the ET, the CC, or the combined ETCC region functionally specified BRD2 and BRD4S. To do this, we devised additional chimeric BET proteins using BRD2 and BRD4S as the backbones onto which the sequences surrounding or containing the ET and CC domains were interchanged (Figure 3.10.A). We can broadly classify these constructs based on the whether they are composed of the BRD2 or BRD4S N-terminal half. For the BRD2-based chimeric constructs, we tested which, if any, of the C-terminal segments (the middle region between BD2 and ET, the ET, or the end segment) resulted in a loss of BRD2 function when the corresponding region from BRD4S was substituted. For the BRD4S-based chimeric constructs, we tested whether the BRD2 ET domain, the CC region or a segment containing both could impart BRD2 function to BRD4S. As controls, we compared these new chimeric constructs to the original N- and C-terminal-swapped chimeras BRD2-4S and BRD4-2. We also compare these constructs to BRD2 and BRD3, which possess the CC region, and to BRD4S, which lacks the CC region (Figure 3.10.B, center). We used a combination of TER119 marker expression and erythroid gene activation to uncover any functional differences between these BET constructs.

We first examined TER119 expression by flow cytometry. Of note, we used low titer viruses that statistically guarantee single-copy gene integration into the cells. Therefore, any differences in protein expression detected by YFP fluorescence were not due to gene copy number. YFP fluorescence revealed that BRD3 and BRD4S were expressed approximately 2-3-fold higher than BRD2 (Figure 3.10.B, left). This difference is less than that observed without the use of YFP fusion (Figure 3.3.E), indicating that the YFP sequence stabilizes BRD2 mRNA or protein to some extent. Next, we examined erythroid maturation based on TER119 surface phenotyping. As expected, a YFP empty vector or no vector infection had no impact on TER119 expression in BRD2 KO cells whereas virus encoding BRD2 restored TER119 close to parental G1E-ER4 levels (Figure 3.10.B, right). BRD3 also restored TER119 expression, although to a lesser extent than BRD2. BRD4S had minimal impact on TER119 expression.

We next examined the relative abilities of the BRD2-based chimeric BET proteins to restore TER119 and erythroid gene expression. All of the BRD2-based constructs are expressed at levels similar to BRD2 based on YFP fluorescence (Figure 3.10.B, left). Replacing the middle region (BRD2-4mid) or the ET domain (BRD2-4ET) from BRD2 with the corresponding regions from BRD4S had no appreciable impact on TER119 expression (Figure 3.10.B, right. However, replacing the end segment of BRD2 (BRD2-4end), which contains the CC region, completely abrogated its activity to the same extent as replacing the entire C-terminal half (BRD2-4S). To examine gene expression, we established stable BRD2 KO1 cells expressing the above constructs at similar levels (Figure 3.11.A, left). Expression of the *Brd2*-based transgenes and control genes (*Actb* and *Gapdh*) was similar across the cell lines (Figure 3.11.A, right). Examination of erythroid gene expression revealed differences in the ability of these constructs to rescue expression of *Hbb-b1*,

Csf2rb, *Csf2rb2* and *Rhag* and to a lesser extent *Alas2*, *Spta1 and Kcnn4* that largely reflect the differences in TER119 expression (Figure 3.11.A, right). For some of these genes, all of the chimeric constructs exhibited reduced activity. However, BRD2-4end had the least activity and most resembled BRD2-4S for 7 of the 8 genes examined. In summary, the activity of the BRD2-based chimeras indicates that the region that functionally discriminates BRD2 and BRD4S is the C-terminal end segment containing the CC domain. The middle segment and the ET domain appear interchangeable with minimal disruption of BRD2 activity and therefore do not contain functionally specific elements.

We next assessed the activities of the BRD4S-based chimeras. The BRD4S-based chimeric BET proteins were expressed at similar levels, about 2-fold higher than BRD2 based on YFP fluorescence (Figure 3.10.B, left). Replacing the BRD4 ET domain with that of BRD2 (BRD4S-2ET) did not increase TER119 expression (Figure 3.10.B, right). Adding the BRD2 CC domain adjacent to the BRD4 ET domain (BRD4S-2CC) likewise did not impact TER119 expression. However, when both the ET and CC domains from BRD2 were inserted into BRD4S (BRD4S-2ETCC), TER119 expression was restored to BRD2-like levels. The BRD4-2 protein, containing the entire C-terminal half of BRD2, had an even greater impact on TER119 expression, indicating that additional C-terminal domains may contribute to BRD2 function. To examine gene expression, we established BRD2 KO1 cell lines stably expressing the above constructs. Protein levels of the BRD4S-based constructs were about two-fold higher than BRD2 protein levels but were similar across the cell lines (Figure 3.11.B, left). Despite the lower protein level of YFP-BRD2, its mRNA was slightly higher than that of the YFP-BRD4S-based BET proteins as measured by qPCR primers targeting *YFP* or the 3' end of *Brd2*, consistent with the decreased protein:RNA ratio previously observed for BRD2 (Figure 3.11.B, right). Levels

of mRNA for *Gata1-ER, Actb* and *Gapdh* were similar. As with TER119 staining, the ET and CC domains were insufficient to confer BRD2-like activity to BRD4S based on the inability of the BRD4S-2ET and BRD4S-2CC proteins to restore erythroid gene activation (Figure 3.11.B, right). However, the combined ETCC domain (BRD4S-2ETCC) clearly imparted BRD2-activity to BRD4S as *Hbb-b1*, *Csf2rb*, *Csf2rb2* and *Kcnn4* expression were at least partially increased above the levels achieved by BRD4S. As with TER119 staining, BRD4-2 construct had the most BRD2-activity at these genes.

To test whether the various BRD4S-based constructs were able to rescue BRD2-dependent hemoglobinization, we visualized cell pellet color after 48 hours of GATA1-ER activation. As seen previously, BRD2 KO1 cells failed to hemoglobinize in comparison to parental G1E-ER4 cells, a defect which is largely rescued by the restoration of BRD2 expression in the BRD2 KO1 cells (Figure 3.11.C). Whereas BRD4S fails to rescue hemoglobinization, substituting the ET and CC regions into the BRD4S backbone partially converts BRD4S into a BRD2-like molecule, but the combined addition of the ETCC region, or the entire BRD2 C-terminal half more fully restores hemoglobinization.

Lastly, we asked whether the ET, CC or combined ETCC domain from BRD2 could confer BRD4S with the ability to promote cell growth. Note that in both parental G1E-ER4 cells and BRD2 KO cells, overexpression of BRD4S inhibits growth, but chimeric BRD4-2, which contains the entire C-terminal half of BRD2, promotes growth when expressed in BRD2 KO cells. To test the role of the ETCC domains, we repeated the cell growth assay after expressing the HA-tagged BRD4S-based chimeric BET constructs in parental and BRD2 KO1 cells. Consistent with our previous experiments, expression of BRD2 promoted cell growth in BRD2 KO1 cells but had no effect in parental G1E-ER4 cells which possess endogenous BRD2 (Figure 3.11.D). BRD4S, on the other hand, repressed parental and

BRD2 KO1 cell growth. Substitution of the BRD2 ET domain into BRD4S (BRD4S-2ET) slightly reduced the growth inhibitory effect of BRD4S. Addition of the BRD2 CC domain adjacent to the BRD4S ET domain (BRD4S-2CC) resulted in a minimal growth promoting effect when expressed in BRD2 KO1 cells. Strikingly, addition of both the ET and CC domains from BRD2 (BRD4S-2ETCC) had the same growth promoting effect as replacing the entire C-terminus with that of BRD2, indicating that the combined ETCC region is the module that functionally distinguishes BRD2 and BRD4S.

Overall, these gene expression, differentiation, and growth assays suggest that neither the bromodomains, the ET domain nor the CC region are alone sufficient to convert BRD4S into a BRD2-like molecule. Instead, the combined region containing the canonical ET domain and the adjacent CC is the critical region encoding BRD2-specific function.

Discussion

We used BET dependent G1E-ER4 cell growth, maturation and gene expression to define functional similarities and differences among BET proteins. We exploited the fact that BRD2 is required for G1E-ER4 growth and GATA1-ER activity. Overexpressed BRD3, but not BRD4S or BRD4L, could significantly overcome defects associated with BRD2 deficiency, pointing to functional similarity among BRD2 and BRD3. When comparing BRD2 and BRD4S, most of their specific activities were determined not by the bromodomains but by the C-terminal halves of the molecules. We identified a short sequence downstream of the ET domain that contains a putative coiled coil (CC region) in BRD2 and BRD3 but not in BRD4. Together with the ET domain, the CC region conferred BRD2-like activity to BRD4S and deletion of the ET or CC domains diminished BRD2 function.

G1E-ER4 cells are a powerful system for the study of BET proteins since they are sensitive to pharmacologic BET inhibition and genetic perturbations of BET proteins[113,139]. Acetylation of GATA1 promotes association with BET proteins[139] which in turn are thought to contribute to GATA1 activity. In this context it is worth noting that the clinical side effects of BET inhibition include anemia and thrombocytopenia[133,134]. Since GATA1 is essential for erythrocyte and megakaryocyte lineage differentiation, it is likely that the detrimental effects of BET inhibition are linked to GATA1 function.

Over the course of our study, we uncovered a discrepancy in the amount of BRD2 protein compared to that of BRD3 and BRD4S when expressed from the same retroviral vector. BRD2 protein was substantially lower even though transcript levels of BRD2 were in fact higher than that of BRD4S, indicating that the mechanism for the low BRD2 abundance is post-transcriptional. The N-terminal half of BRD2 was responsible for the low protein amounts, as the protein levels of the chimeric construct BRD2-4S were similar to BRD2 whereas the protein levels of the chimeric construct BRD4-2 were similar to BRD4S. By fusing YFP to the N-terminus of BRD2, the discrepancy between BRD2 protein and mRNA was reduced, though YFP-BRD2 protein was still lower than that of YFP-BRD3 and YFP-BRD4S. The YFP portion might thus affect BRD2 protein production or turnover. It is interesting to note that a similar discordance between overexpressed BRD2 and BRD3 protein was observed in U2OS cells but not HEK293T cells[31]. Cell-type specific mechanisms may exist to tightly regulate BRD2 protein levels. SPOP-mediated proteasomal targeting has been implicating in

regulating BET proteins in some models[156–158]. However, SPOP recognizes a degron motif shared equally by each BET protein. Thus, this mechanism is unlikely to account for the differences we observed between BRD2/3 and BRD4S in G1E-ER4 cells. Additional studies are therefore required to elucidate what accounts for this specificity.

Our unexpected finding that the BRD2 bromodomains can be exchanged for those of BRD4 with no substantial loss of BRD2 function suggests that potential differences in their affinities or specificities for acetylated substrates[44,47,48,51–53] do not seem to contribute much to their selective functions. Our results suggest that the distinct functions of BRD2 and BRD4 are instead exerted by their C-terminal halves, in which the extended ET region containing the CC in BRD2 but not BRD4 appears to be the critical element. Gene rescue experiments showed that ET domain itself was essential for BRD2 function, which is consistent with previous studies demonstrating that the ET domain recruits multiple chromatin regulators including NSD3, CHD8, ATAD5 and JMJD6[31,64,66,68,71,72]. However, in our study, though the ET domain was essential, exchanging the BRD2 and BRD4S ET domains did not change the activities of either protein. Instead, the BRD2 ET domain appears to function in conjunction with the adjacent CC. Thus, when the combined ETCC region was grafted into BRD4S, it imparted BRD2-like activity to BRD4S.

Coiled coils are structural motifs present in a wide variety of proteins and perform a multitude of functions including directly mediating protein-protein interactions[159]. In this case, the ET proximal CC may be recruiting additional

factors that are ultimately altering the function of BRD2 in comparison to BRD4. Alternatively, the CC domain may be modulating the structure of the ET domain by influencing its folding to create novel interaction surfaces at the junction. Protein interaction and structural studies are therefore warranted to better understand how this region contributes to BRD2 functional specificity.

The ETCC region may therefore allow individual BET proteins to incorporate into different transcriptional complexes involved in discrete aspects of the transcription cycle. This may at least in part reconcile how BET proteins can function distinctly on chromatin while having relatively similar genome-wide distribution patterns, as observed in many studies[55,63,96,102,114,117,118,129,160]. Protein interactions mediated by the ET or ETCC regions may also account for bromodomain-independent chromatin binding[36]. Thus, the complexes that each BET protein associate with could also account for some of the differences in chromatin occupancy exhibited by each BET protein, as is the case at enhancers where there is a relative enrichment BRD4 and at promoters where there is a relative enrichment of BRD2 and BRD3[113,117,129]. There is also evidence that the conserved mB domain, which facilitates homo- and hetero-dimerization of BET proteins, is required for chromatin binding[35]. Heterodimerization of BET proteins might further explain their overlapping chromatin occupancy and may suggest that BET proteins function coordinately to assemble multiple factors onto chromatin. Together these considerations lead to a model in which BET proteins function both independently and interdependently during transcription.

At low doses of BET inhibitors, the expression of only a subset of genes is affected in a manner that is not easily predictable based solely on local chromatin features or levels of BET protein occupancy[112,113]. Predicting whether a gene responds to pharmacologic BET inhibition remains a challenge in part because cells express multiple BET proteins exhibiting mostly overlapping but partially distinct chromatin occupancy patterns. Moreover, partial redundancy among BET proteins, as exemplified by BRD2 and BRD3, might be rooted in structural and functional similarities. Thus, overexpression of BRD3 at least partially compensates for loss of BRD2[113]. This example illustrates that different BET protein dependencies can also be a result of the expression levels of each protein. These factors need to be taken into consideration when assessing the effects of targeting BET proteins pharmacologically.

Hence, identifying BET protein specific mechanisms and predicting what disease processes might be targeted by BET inhibitors remains an important goal. Achieving this level of mechanistic understanding requires consideration of BET protein expression levels, assessment of functional overlap among BET proteins, and the identification of functional domains that are required for BET protein activity in diverse contexts. Our work begins to address these criteria by identifying the ETCC module, which may mechanistically differentiate BRD2, and given the structural similarities also BRD3, from BRD4. In the next Chapter, we explore the mechanisms by which this domain may contribute to transcription. We speculate that targeting this region with a small-molecule inhibitor might provide a more selective tool for therapeutic BET inhibition.

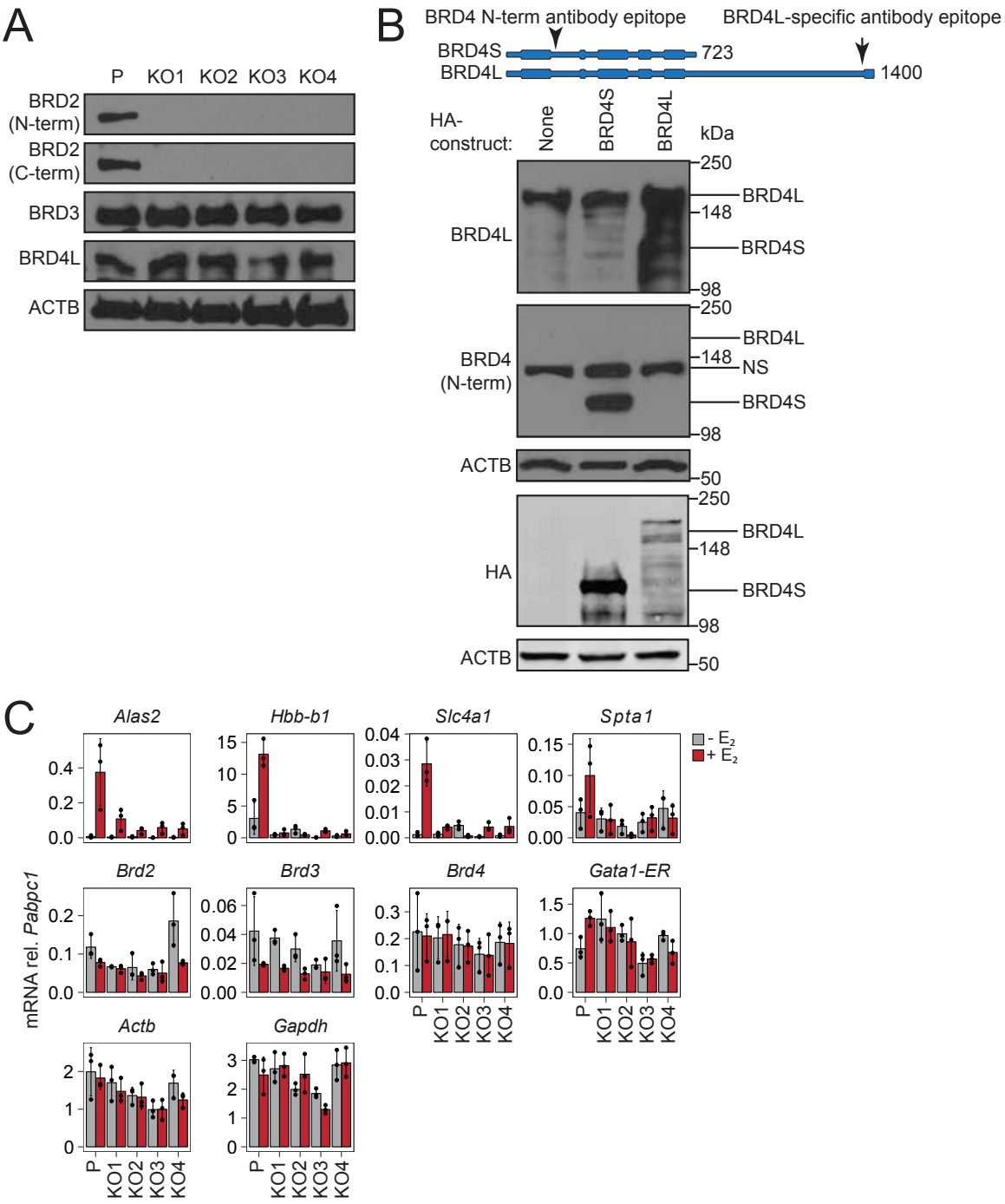


Figure 3.1. *Brd2*-null G1E-ER4 cells fail to activate erythroid genes. A) Western blot of endogenous BET protein levels in whole cell lysates from parental G1E-ER4 cells or multiple *Brd2*-null clones (KO1-KO4). B) Western blot of endogenous and overexpressed BRD4 isoforms using an N-terminal antibody that is predicted to detect both BRD4S and BRD4L and a C-terminal antibody that only recognizes BRD4L. Whole cell lysates from parental G1E-ER4 cells expressing no construct or HA-tagged BRD4S or BRD4L. C) Gene expression levels normalized to *Pabpc1* after 24 hours of GATA1-ER activity (+E_2)

in the indicated cell lines. Averages and standard deviation (error bars) are derived from three independent replicates.

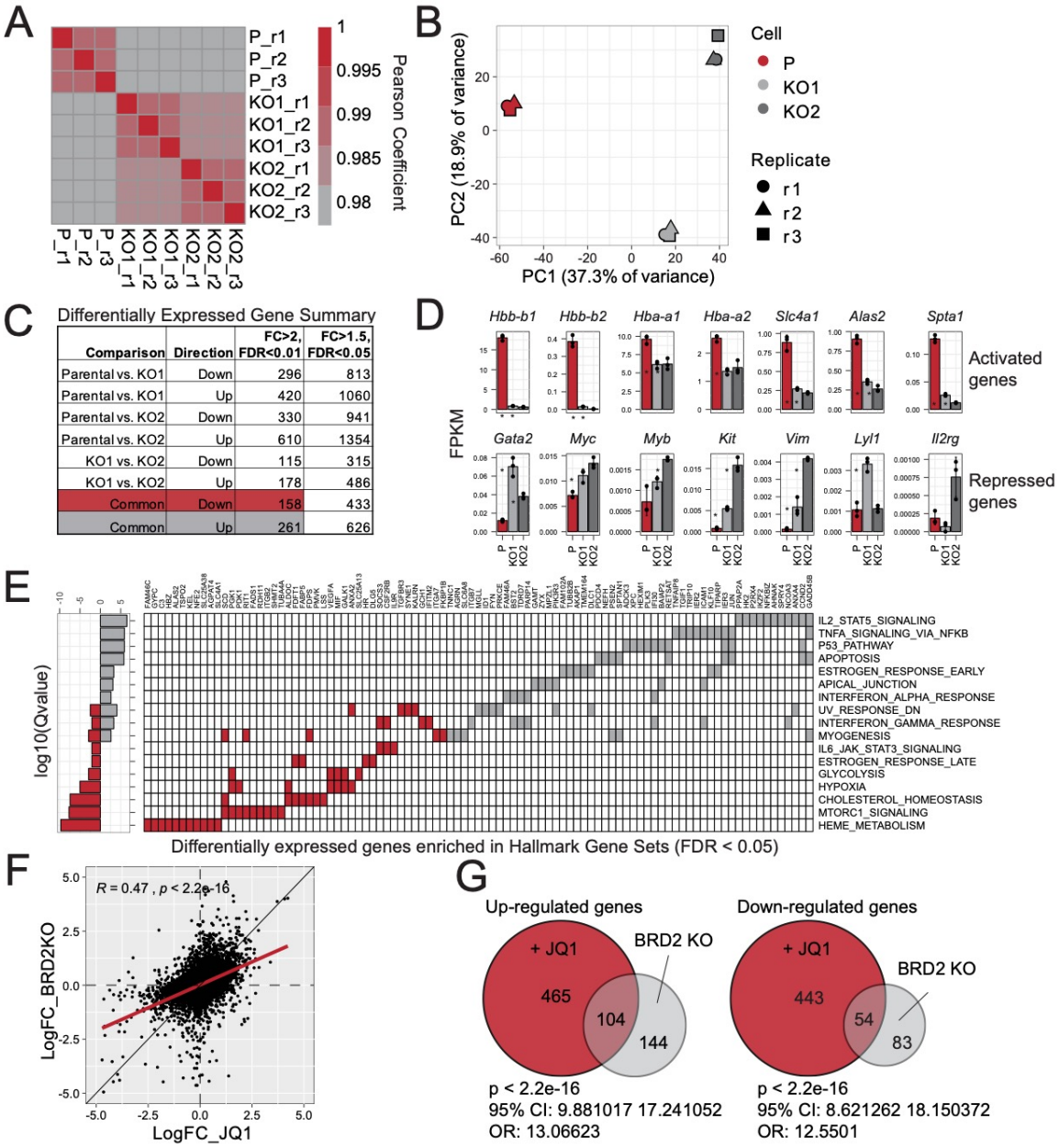


Figure 3.2. BRD2-dependent gene expression. A) Heatmap depicting Pearson coefficients of pairwise gene expression correlations in three independent RNAseq replicates (r1-r3) after GATA1 activation. B) Principle component analysis (see text for details). C) Summary of differentially expressed genes (FC=fold-change, FDR=false discovery rate). Subsets used for subequent analysis shaded in red or gray. D) Normalized read counts (FPKM) for indicated genes activated or repressed upon GATA1-ER activation. Averages and standard deviation (error bars) are derived from three independent replicates. E) Enrichment scores (log10Qvalue, left) and gene names (right) in the top ten Hallmark gene sets from the Molecular Signatures Database enriched in the up- and down-regulated genes. F-G) Correlation of all detected genes (F) and intersection of most differentially expressed (FC>2, FDR<0.01) genes (G) after 24 hours of JQ1

treatment or BRD2 depletion in GATA1-ER activated cells. P value (p), 95% confidence interval (CI), and odds ratio (OR) of overlap calculated by Fisher's Exact Test.

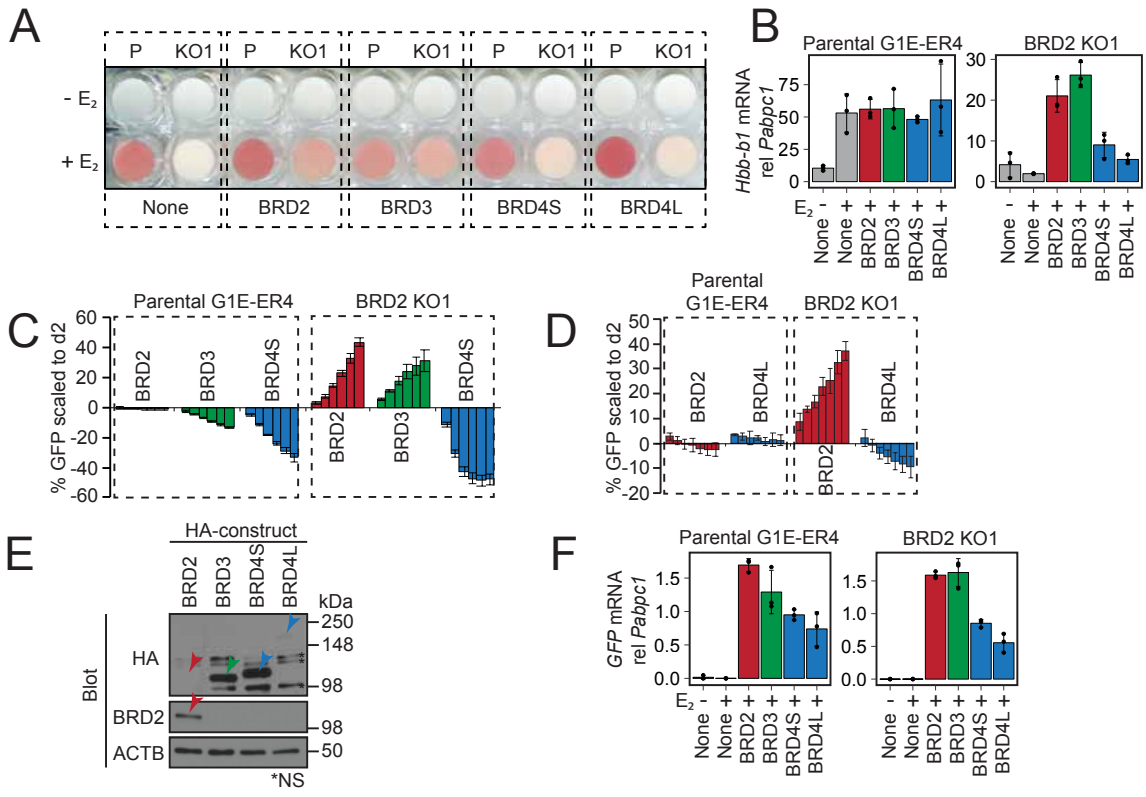


Figure 3.3. Differential function and expression of BRD2, BRD3 and short and long isoforms of BRD4 in G1E-ER4 cells. A) Visualization of hemoglobin production (redness) with or without 48 hours of GATA1-ER activation (+E_2) in parental (P) or *Brd2*-null (KO1) G1E-ER4 cells stably expressing the indicated HA-tagged BET proteins after retroviral transduction. B) *Hbb-b1* mRNA levels normalized to *Pabpc1* after 24 hours of GATA1-ER activation. C-D) Competitive growth assay in which %GFP is tracked after retroviral transduction with bicistronic vector encoding indicated BET protein and GFP selection marker. Bars indicate consecutive days of GFP percentage measurement from left to right starting with day 2 (d2) and ending with d8. Starting %GFP has been subtracted to scale measurements to same axis. E) Western blots of HA-tagged BET protein expression in whole cell lysates from BRD2 KO cells, representative of similar expression patterns seen in lysates from parental G1E-ER4 cells. *NS = non-specific bands associated with HA antibody. F) GFP mRNA levels (indicative of transgene expression) normalized to *Pabpc1*. Averages and standard deviation (error bars) are derived from three independent replicates.

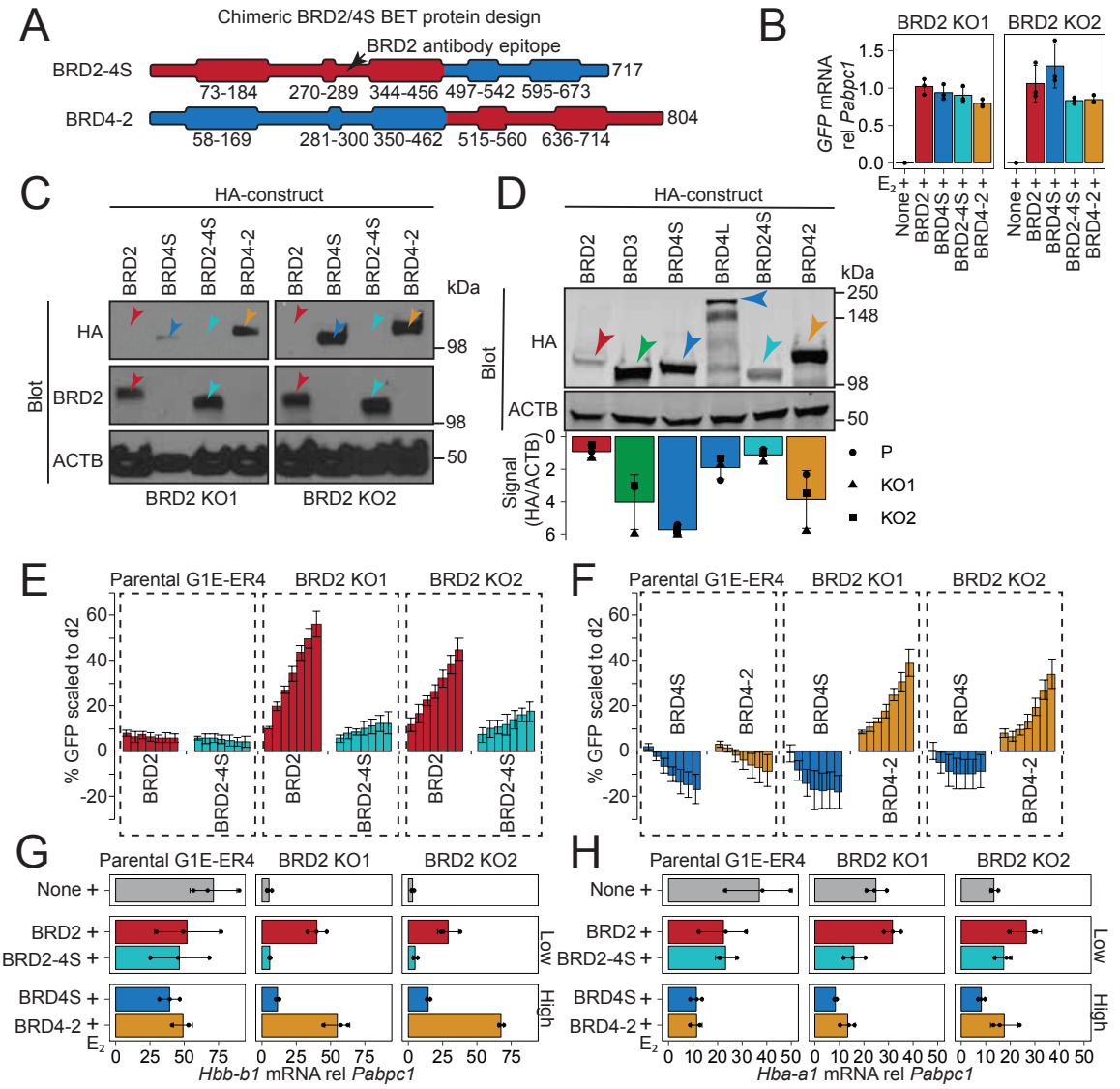


Figure 3.4. Structure, expression and function of chimeric BET proteins composed of N- and C-terminal halves of BRD2 and BRD4S. A) Design of chimeric BET proteins Location of epitope recognized by BRD2 antibody used in (C) is indicated. B) GFP mRNA levels (indicative of transgene expression) normalized to *Pabpc1* after 24 hours of GATA1-ER activation (+E_2). C) Western blot of HA-tagged BET protein expression in whole cell lysates from BRD2 KO cells, representative of patterns also observed in lysates from parental cells. NS = non-specific bands associated with HA antibody. D) Quantitative western blot comparing protein expression levels of chimeric BET proteins with intact BET proteins in BRD2 KO1 whole cell lysates, representative of expression patterns in parental and BRD2 KO2 lysates. ACTB normalized signal plotted below. E-F) Competitive growth assay in which %GFP is tracked after retroviral transduction with bicistronic vector encoding indicated BET protein and GFP selection marker. Bars indicate consecutive days of GFP percentage measurement from left to right starting with day 2 (d2) and ending with d8. Starting %GFP has been subtracted to scale measurements to same axis. G-H)

Hbb-b1 and *Hba-a1* mRNA levels normalized to *Pabpc1* after 24 hours of GATA1-ER activation. Constructs grouped by expression levels. Averages and standard deviation (error bars) are derived from at least three independent replicates.

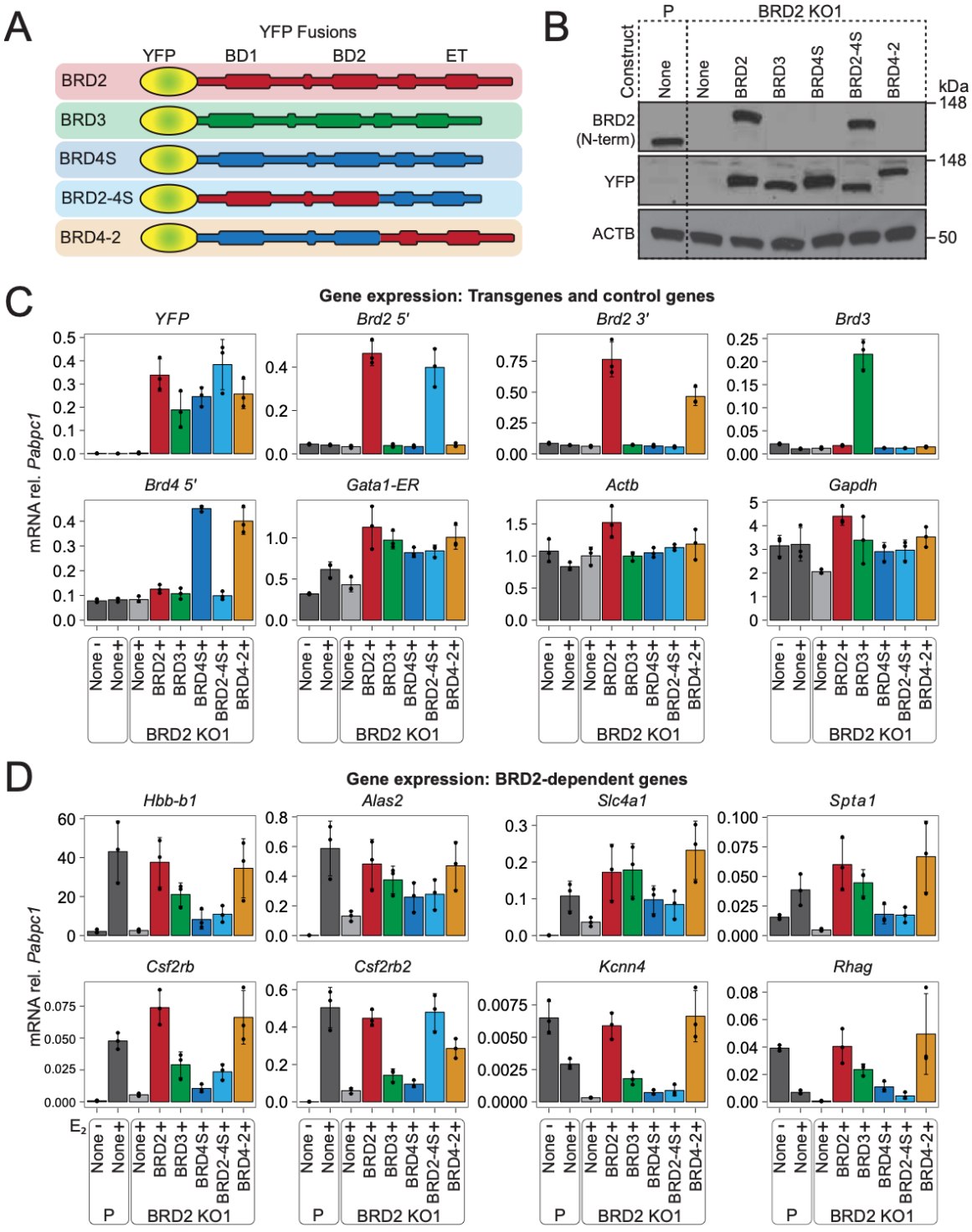


Figure 3.5. The C-terminal halves of BRD2 and BRD4S distinguish their function. A) Schematic of YFP-fusion constructs utilized in this experiment. B) Western blot detection of exogenous YFP-fused BET proteins in whole cell lysates after expression in BRD2 KO1 cells with detection of endogenous BRD2 in parental G1E-ER4 cell for comparison. C)

Gene expression levels of BRD2-dependent and control genes normalized to *Pabpc1* after 24 hours of GATA1-ER activity (+E_2). 5' and 3' indicate the respective halves of the mRNAs. Averages and standard deviation (error bars) are derived from three independent replicates.

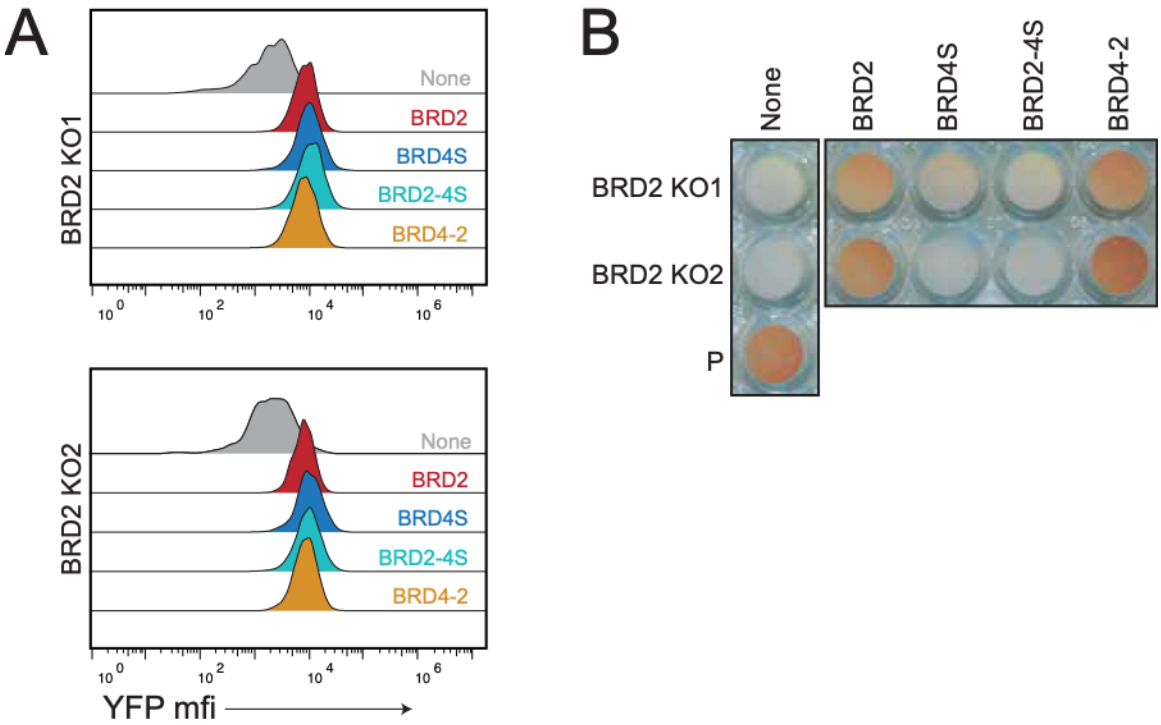


Figure 3.6. Control gene expression and hemoglobin synthesis in BRD2 KO cells expressing YFP-BET proteins. A) YFP-BET protein levels after iterative cell sorting measured by flow cytometry. B) Visualization of hemoglobin production (redness) after 48 hours of GATA1-ER activation.

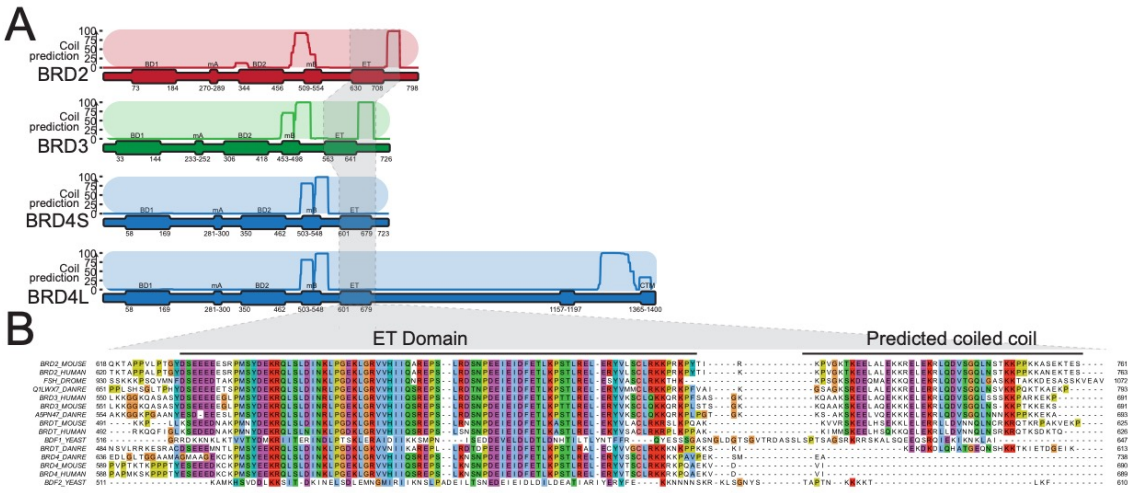


Figure 3.7. Identification of a conserved putative coiled coil (CC) domain adjacent to the ET domain. A) MARCOIL coiled coil prediction across the mouse BRD2, BRD3, BRD4S and BRD4L amino acid sequences. B) Conservation of the ET and predicted coiled coil region between BET proteins and across species. Color coding: blue (hydrophobic), red (positive charge), magenta (negative charge), green (polar), pink (cysteines), orange (glycines), yellow (prolines), cyan (aromatic) and white (not conserved).

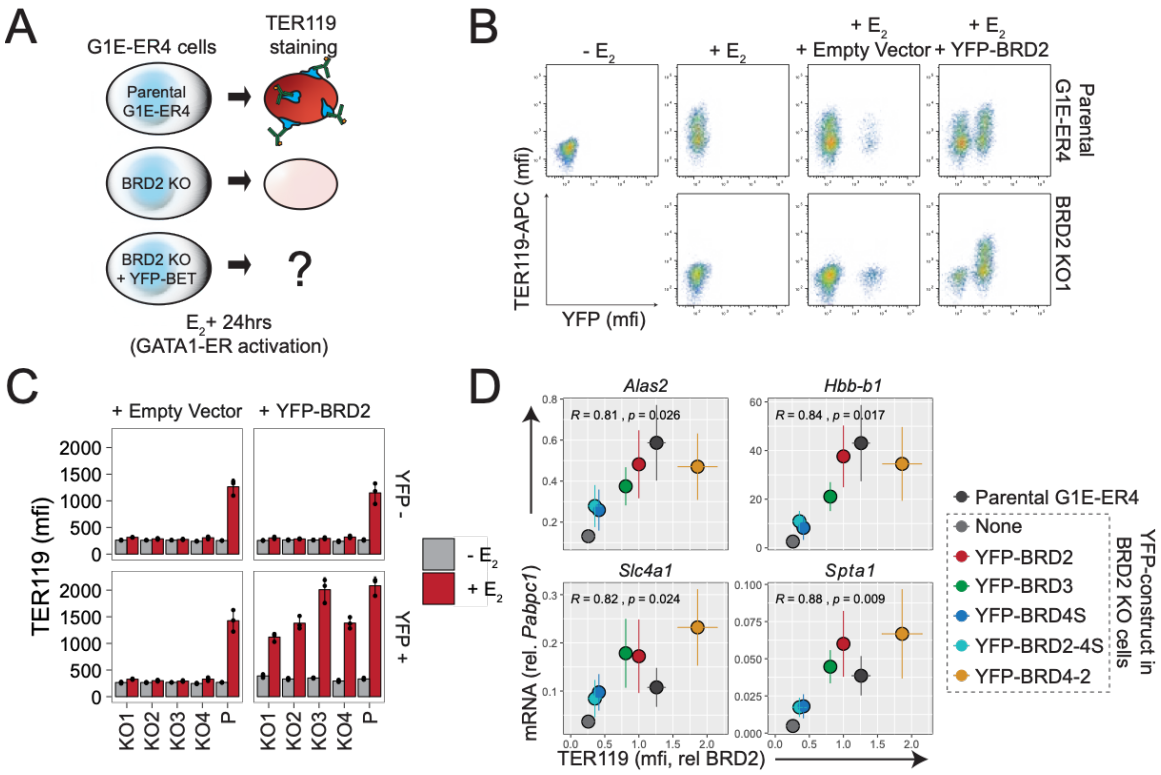


Figure 3.8. Validation of TER119 erythroid maturation assay. A) Schematic depicting TER119 detection after 24 hours of GATA1-ER activation (+E2). B) Representative flow cytometry data of parental G1E-ER4 cells or BRD2 KO1 cells expressing an empty YFP vector or YFP-BRD2 with or without 24 hours of GATA1-ER activation (+E2). C) Summary data quantifying mean fluorescence intensity (mfi) of TER119 staining from three independent replicates. D) Scatter plots depicting relationship between the amount of TER119 surface expression for indicated cell lines and the amount of mRNA for the indicated erythroid genes upon 24 hours of GATA1-ER activity. R = Pearson coefficient. Averages and standard deviation (error bars) are derived from three independent replicates.

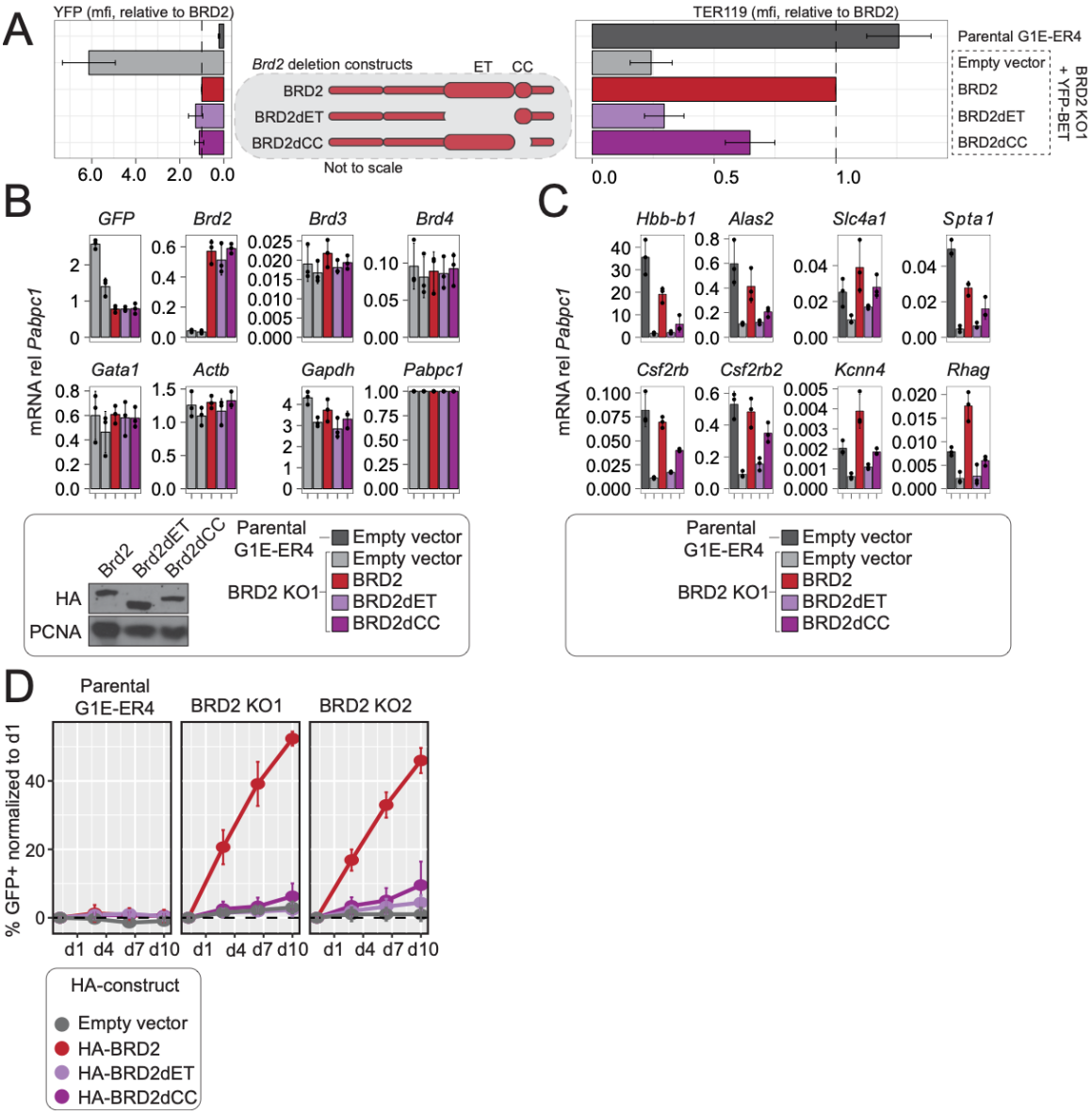


Figure 3.9. The ET and CC domains each contribute to BRD2 function. A) Schematic of BRD2 constructs with the ET and CC domains deleted (middle) and their resultant protein expression levels (left) and effect on TER119 surface expression (right) when expressed in three independent BRD2 KO cell clones after 24 hours of GATA1-ER activation. Values are normalized to YFP-BRD2. B-C) Gene expression levels of control genes (B) and BRD2-dependent genes (C) normalized to *Pabpc1* in cells stably expressing the indicated HA-tagged BET proteins after 24 hours of GATA1-ER activation. Inset. Western blot of HA-tagged BET protein levels. D) Competitive growth assay in which %GFP is tracked after retroviral transduction with bicistronic vector encoding indicated BET protein and GFP selection marker. Starting %GFP has been subtracted to scale

measurements to same axis. Averages and standard deviation (error bars) are derived from three independent replicates.

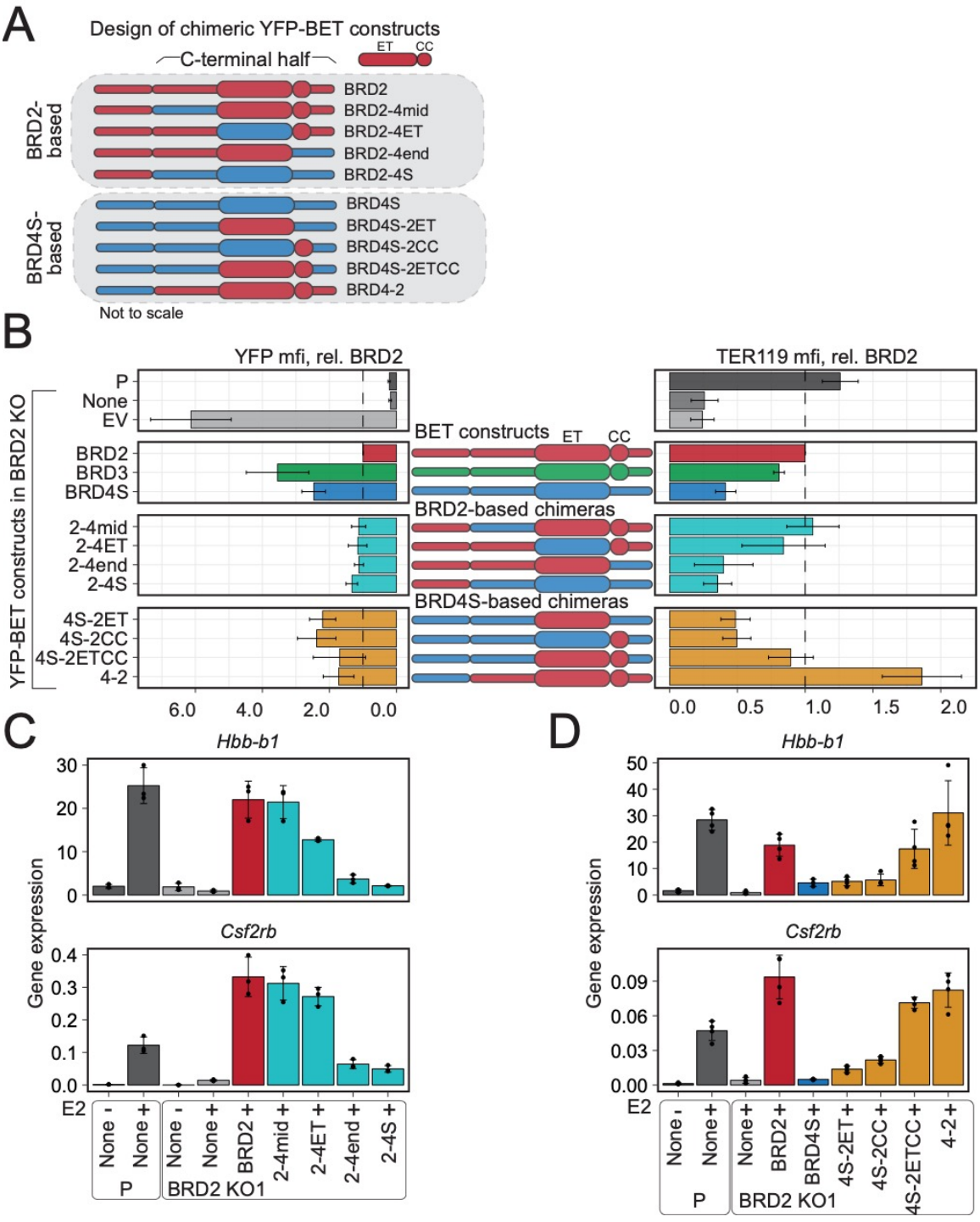


Figure 3.10. THE ETCC region functionally distinguishes BRD2 and BRD4S. A) Schematic of the YFP-BET fusion constructs in which C-terminal regions have been interchanged. B) Protein expression levels (left) and TER119 surface levels (right) of YFP-BET constructs after 24 hours of GATA1-ER activation. Values are normalized to YFP-BRD2. C-D) Expression levels of mRNA normalized to *Pabpc1*

after 24 hours of GATA1-ER activity (+E2). Averages and standard deviation (error bars) are derived from three to four independent replicates.

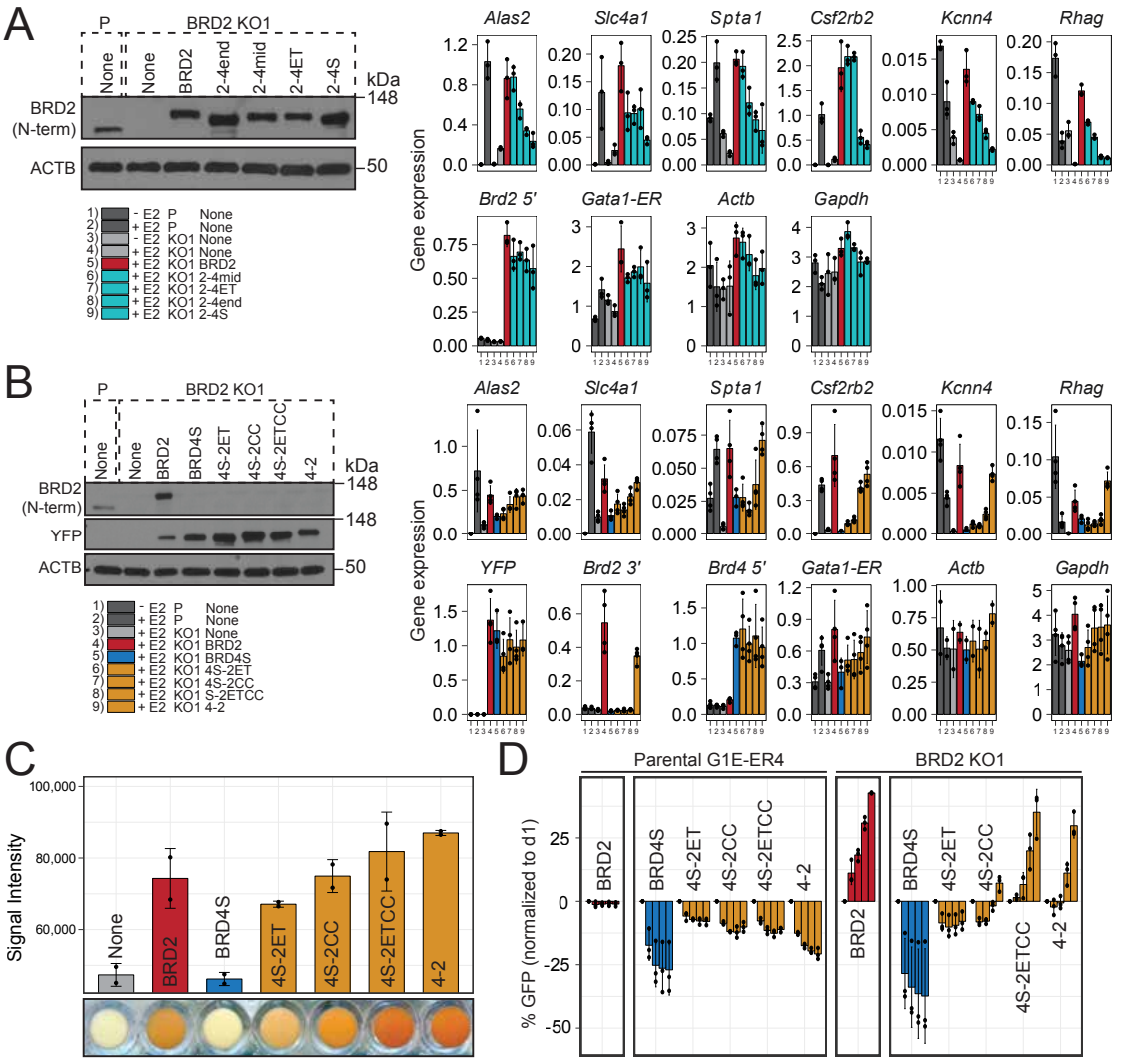


Figure 3.11. Protein levels, gene expression, hemoglobinization, and growth associated with chimeric BET proteins. A-B) Left: western blots of YFP-BET proteins in G1E-ER4 (P) or BRD2 KO1 lysates. Right: gene expression of BRD2-dependent genes (top row) and control genes (bottom row) normalized to *Pabpc1*.C) Hemoglobinization of BRD2 KO1 cells with total color signal intensity averages across two replicates (error bars depict standard deviation). D) Competitive growth assay in which %GFP is tracked after retroviral transduction with bicistronic vector encoding indicated HA-tagged BET protein and GFP selection marker. HA-BRD2-expressing cells have been grouped separately based on lower protein expresson levels (data not shown). Bars indicate subsequent days of GFP percentage measurement (days 1, 4, 7, 10, 13).

Genes down-regulated in *Brd2*-null G1E-ER4 compared to parental cells after 24 hours of GATA1-ER activation						
Alas2	Hbb-b1	Slc4a1	Ctsz	Morc2a	Fdps	Sdsl
Aldoc	Hr	Tgfbr3	Mrpl54	Fam46c	Slc25a38	Arhgef40
Bnip3	Il9r	Tgm1	Cystm1	Fads1	Syne3	Shc4
Btk	Itga7	Trap1a	Smim5	Srxn1	Mylk3	Hist1h2af
C3	Itgb2	Tuba4a	Grtp1	Nxpe2	Slc25a21	C5ar2
Anxa2	Kcnb1	Vegfa	Cntd1	Ifitm2	Plekhb2	Tcp11l1
Cetn3	Kcnn4	Slc23a3	Wfdc15a	Trim2	Fam132b	Treml2
Socs3	Fabp5	Kel	Agpat4	Sec16b	Pdk1	Il1bos
Coq7	Lrp8	Gcat	Pmvk	Igsf9	Ctso	Rph3al
Csf2rb	Lss	Hs6st1	Cd177	Hemgn	Phf24	Tmem29
Csf2rb2	Rdh11	Slc25a13	Zg16	Hecw1	Oscp1	Fam205a1
Esd	Mif	St3gal6	Pdia2	Igflr1	Pla2g4c	Ccdc8
Foxh1	Ncf4	Higd1a	Slc52a3	Ttc41	Zfp939	Hbq1b
Fdft1	Nfe2	Ctsf	Ptrhd1	Tcrg-C1	Zfp599	Kalrn
Fkbp1b	Pctp	Mocs1	Tspo2	Scml2	Ttc12	Tubb1
Slc6a13	Pgk1	Dnajb2	Dlg5	Shmt2	Pcyt1b	Myh7b
Gch1	Rit1	Sult5a1	Gypc	Ston2	Zfp870	Hbb-bs
Arhgdig	Scd2	Slc43a3	Esco2	Ank2	Plekhm3	
Galk1	Spi1	Arhgap23	Cyp2c55	Hyal3	Ypel4	
Hba-x	Six4	Syne1	Uba7	Arrb1	C2cd4a	

Table 3.1. Gene expression analysis of *Brd2*-null G1E-ER4 cells: down-regulated genes. Differentially expressed genes in *Brd2*-null G1E-ER4 cells were identified based on >2-fold change in expression compared to parental cells after 24 hours of GATA1-ER activation. Genes were only included if identified in both BRD2 KO1 and KO2 cell lines.

Genes up-regulated in Brd2-null G1E-ER4 compared to parental cells after 24 hours of GATA1-ER activation								
Abca2	Ier2	Rom1	Map3k14	Ydjc	Pag1	Znrf1	Snx33	Nckap5l
Agrn	Ier3	Pmel	Sept9	Bst2	Cnnm4	Il17rc	Gk5	Pdp1
Akap1	Itgb7	Slc19a1	Cdkl2	Mfsd3	Sfxn3	Fut10	Nbeal2	Zfp69
Akr1b3	Itpr1	Sms	Rfx5	Fastkd3	Fam102a	Galnt10	Irf2bpl	Nxpe3
Alox5	Itpr2	Sntb2	Trim3	Cpne3	Tiparp	Mical1	Ccnb1ip1	Gnptab
Slc25a4	Itpr3	Serpinb9	Espn	Slc29a3	Tdrd7	Hexim1	Exd1	Acsl5
Anxa4	Jun	Sptan1	Sept6	Cotl1	Pafah2	Pla2g15	Usp35	Parp14
Aprt	Kcnu1	Srm	Rec8	Zfp661	Acacb	Grhl1	Peak1	Stard9
Rnd2	Khk	Srpk2	Bcam	Lrrc8e	Slc6a8	Gramd1c	Fam199x	Fam129c
Cbfb	Kifc2	Stk10	Kcne3	Fgfbp3	Etv5	Ccdc62	Acsf3	
Ccnd2	Arhgef2	Stx1a	Extl2	Mblac2	Sbk1	Thnsl1	Pkn3	
Entpd6	Lcat	Tac2	Dip2a	Top1mt	Srl	Zfp710	Bahcc1	
Plk3	Lgals9	Tep1	Dip2a	Tmem192	Ppcs	Tmem164	Xxylt1	
Ctsl	Limk1	Tgif1	Ifi30	Ccdc103	Trip10	Suox	Znf512b	
Dlx1	Man2b1	Klf10	Asph	Itpripl1	Tnfaip8	Fam46a	Irf2bp2	
Dnmt3b	Mfge8	Tle2	Lyrm9	Tubb2b	Osr2	Slc35e3	Maml2	
Lefty1	Map4	Tnnc1	Ahnak	Iqce	Frs3	Ccdc92	Mbtps2	
Nr2f6	Gadd45b	Tnni3	Rnf128	Ppapdc2	Baiap2	Pdxk	Tango6	
Fmo5	Myo5a	Tuft1	Bzw2	Pomk	Nabp1	Kctd11	Tmcc3	
Fyn	Ncoa3	Ugdh	Pqlc1	Cklf	Fam212b	Plekhh3	Ccdc171	
Fzd7	Nnt	Xpc	Ptgr1	Exoc6b	Ssc4d	Cdca7l	Bcorl1	
Gamt	Nnt	Ikzf2	Fam132a	Lhpp	Pygb	Tmem184b	Itpkb	
Gata2	Oprl1	Zyx	Adck3	Ppp1r18	Ggct	Rnd1	Gchfr	
Gata3	P2rx4	Mgll	Retsat	Klhl2	Amhr2	Camk1d	Soga1	
Gip	Pdcd4	Spry4	Ppfibp1	Prss36	Nr3c2	Tnks1bp1	Nol4l	
Hhex	Pik3r3	Homer2	Prpsap1	Eml4	Kiss1r	Cyb5rl	Adgrl1	
Hk2	Prkce	Dlc1	Dynll2	Zfp623	Dok4	Zfp382	Piwil4	
Hoxc8	Ppap2a	Galns	Mpzl1	Srd5a1	Tle6	Zfp764	Acad12	
Icam1	Pros1	Hgsnat	Gpr155	Nfkbiz	Man2a2	Fhod1	Muc6	
Id1	Psen2	Nucb2	Zfp707	Kremen1	Igf2bp1	Klhl36	Nefh	

Table 3.2. Gene expression analysis of *Brd2*-null G1E-ER4 cells: up-regulated genes. Differentially expressed genes in *Brd2*-null G1E-ER4 cells were identified based on >2-fold change in expression compared to parental cells after 24 hours of GATA1-ER activation. Genes were only included if identified in both BRD2 KO1 and KO2 cell lines.

CHAPTER 4: THE COILED COIL CONTRIBUTES TO UNIQUE PROTEIN INTERACTIONS

Chapter summary

Research described in this chapter is a combination of experiments that will be included in a pending publication "Comparative structure-function analysis of bromodomain and extraterminal motif (BET) proteins using a gene-complementation system" and unpublished work. This chapter will include unpublished chromatin immunoprecipitation (ChIP) sequencing data that was generated in collaboration with Cheryl Keller and Belinda Giardine in Ross Hardison's laboratory at Pennsylvania State University. The analysis of this data includes comparisons to previously published data generated by Aaron Stonestrom and additionally analyzed by Sarah Hsu while they were in the Blobel lab[113,131]. This chapter also includes proteomic data that will be published in the above manuscript. Hongxin Wang was critical to the generation of the proteomic data and subsequent validation experiments. The mass spectrometry data was conducted in collaboration with Hsin-Yao Tang and Thomas Beer in the Proteomics and Metabolomics Facility at the Wistar Institute. Perry Evans of the CHOP Department of Biomedical and Health Informatics assisted with analysis of this data. Structural characterization of the coiled coil region of BRD2 and BRD3 was performed by Yichen Zhong and Joel Mackay at the University of Sydney.

Experiments examining the interaction between BRD2 and PAF/CK2 in more detail remain unpublished and will be the basis of future work to determine the functional significance of these findings.

Introduction

BET proteins are being investigated as therapeutic targets in a number of solid and hematological cancers. However, treating these diseases with anti-BET therapy will require an intimate knowledge of how BET proteins regulate transcription. A simplistic model for BET protein function begins with the observation that BET bromodomains form contacts with acetylated lysine residues on active chromatin. Various chromatin regulators that influence POL2 activity are then recruited through protein-protein interactions. Because BET inhibitors typically disrupt the contact between the bromodomains and their acetylated substrates, the mechanism of action of these drugs is thought to be the displacement of POL2 regulators from chromatin.

The exact details of this model are more difficult to discern. For example, though it is appreciated that some BET proteins (particularly BRD2 and BRD4) can function distinctly in their roles as transcriptional regulators, it remains unclear how this is achieved mechanistically. One hypothesis is that individual BET proteins participate in distinct protein interactions. This is supported by immunoprecipitation experiments demonstrating differential enrichment of protein complexes by each BET protein. For example, interactions with the negative elongation factor (NELF) and PTEFb are dominated by BRD4, whereas interactions with casein kinase II (CK2) and TFIID, a subunit of the POL2 complex, are preferred by BRD2 and BRD3[31]. With the exception of the contacts made between BRD4 and PTEFb[76], a mechanistic basis for these preferential interactions remains unclear. For example, CK2 is known to phosphorylate BET proteins at sequences that are well conserved

across BET proteins[85], which would suggest equivalent interaction frequencies, not protein-specific ones. It is possible that BET proteins possess additional interaction domains that specify some of these binding events. One potentially significant domain is the ET domain. This domain has been well established as protein-protein interaction region. Its crystal structure has been solved and a consensus binding motif has been identified in many of its validated substrates[66,68,72]. These substrates include many known chromatin regulators such as ATAD5, JMJD6, GLTSCR1, WHSC1L1/NSD3, and NIPBL. Because the ET domain is approximately 80% conserved across BET proteins, it remains to be determined whether the remaining variation may confer some of the BET protein-specific interactions that have been observed.

The interaction between BRD4 and PTEFb (composed of the CDK9 and CCNT1) is perhaps one of the best studied molecular mechanisms involving BET proteins. The specificity of PTEFb for BRD4 (and for BRDT in the testis) is attributed to the CTM domain[76]. It is thought that BRD4 recruits PTEFb which phosphorylates POL2 and releases it from its paused state. However, several observations call into question the specificity and functional significance of this mechanism. First, though this interaction is strongest for BRD4, BRD2[122] and BRD3[30] have also been shown to immunoprecipitate with CDK9. This may be driven in part by the ability of the ET domain[76] and the second bromodomain[161] to contribute to the PTEFb interaction. Second, regardless of which BET proteins recruit PTEFb, rapid depletion of BET proteins inhibits POL2 elongation without altering PTEFb levels on chromatin[111], indicating an elongation role independent

from PTEFb activity. Accordingly, BET inhibitors and CDK9 inhibitors impact different subsets of genes[112]. In fact, in vitro transcriptional assays that do not include PTEFb components reveal that all BET proteins exhibit histone chaperone-like activity (defined as the ability to facilitate passage of POL2 through nucleosome templates)[40,41]. In the case of BRD4, this activity is independent of its CTM domain. Whether this aspect of BET protein function accounts for all transcriptional defects associated with BET inhibition remains unclear. It is also unclear if BET proteins use similar or distinct mechanisms to carry out this activity.

Another hypothesis regarding distinct BET protein functions stems from the observation that BET proteins can bind to unique regions of the genome. For example, though there is a large degree of overlap, BRD2 and BRD3 bind more predominately to gene promoters, whereas BRD4 favors enhancers. The basis for these patterns is not known, but it is possible that the preference exhibited by BRD2 and BRD3 for promoters is related to their interactions with the histone variant H2A.Z[57–59]. We and others have reported additional BRD2 and BRD3 binding at CTCF sites, where they are thought to cooperate with CTCF to maintain TAD boundaries[131] and regulate enhancer-promoter looping[122]. Notably, BRD4 occupancy at these sites is minimal. This finding is consistent with previous work from our laboratory demonstrating that BRD2 and BRD3 exhibit overlapping functionality that is distinct from BRD4[113]. It is possible that these critical differences in genomic occupancy also underpin the unique transcription signatures apparent upon BRD2 and BRD4 depletion observed in other studies

and perhaps even the different developmental phenotypes associated with *Brd2*- and *Brd4*-null mice.

In Chapter 3, we sought to identify the domain that biologically distinguishes BRD2 and BRD3 from BRD4. We found this domain to encompass the canonical ET sequence shared by BRD2/3/4 and an adjacent coiled coil found in BRD2/3 but not BRD4. We termed this BRD2/3-specific region the ETCC domain. In this chapter, we examine the genomic distributions of YFP-tagged BRD2/3/4S and test whether the C-terminal halves, either containing or missing the newly defined ETCC region, control differential BET protein occupancy. Given the proximity to the ET protein interaction domain, we then examine whether the ETCC mediates different protein-protein interactions.

YFP-BET proteins share similar genome-wide binding profiles

BET proteins bind to similar regions of the genome – particularly promoters and enhancers – and largely colocalize genome-wide. However, some exceptions have been identified. For example, we and others observed BRD2/3-specific binding at CTCF sites. To determine whether the distinct functions of YFP-tagged BRD2, BRD3, BRD4S, and the chimeric BRD2/4S constructs are related to these different binding profiles, we performed chromatin immunoprecipitation followed by deep sequencing (ChIP sequencing) in the BRD2 KO1 cells expressing the YFP-tagged BET proteins. Western blots detecting the levels of YFP-BET fusion proteins in these cells are in Figure 2.5. Note that the levels of ectopic YFP-BET protein are roughly similar to the level of endogenous BRD2 found in parental cells, minimizing the risk of overexpression artifact. We use the YFP moiety shared by

all of the ectopic BET proteins as the epitope for immunoprecipitation, allowing the use of the same antibody to enrich chromatin associated with each BET isoform. These technical considerations minimize biases related to protein expression and antibody affinity/specificity.

We find strong reproducibility between replicates when assessing ChIP sequencing signal at gene transcription start sites (TSS) and previously annotated enhancer elements[162] with Pearson correlation coefficients for each pair of replicates ranging from 0.9036 to 0.9725 at TSSs and from 0.8572 to 0.9349 at enhancers (Figure 4.1.A, Table 4.1). We therefore used the average of the binned read counts for each technical replicate in subsequent analyses. We next asked whether the YFP-tagged BET proteins were differentially bound at GATA1 and CTCF sites using previously published GATA1 and CTCF peaks in activated GATA1-ER4 cells[113,131]. All five YFP-BET proteins were enriched at GATA1 sites, though the average levels of YFP-BRD4S and BRD2-4S were slightly lower than those of YFP-BRD2, BRD3, and BRD4-2 (Figure 4.1B). Surprisingly, none of the YFP-BET proteins were enriched at CTCF sites (Figure 4.1B). Because we had previously observed colocalization of BRD2 and CTCF[131], we repeated endogenous BRD2 ChIP sequencing in G1E-ER4 cells. We observed strong BRD2 signal overall but found no BRD2 occupancy at CTCF sites (Marit Vermunt, personal communication [data not shown]). This discrepancy could be due to technical differences between the studies. For example, the previous BRD2 antibody manufacturer lot number was unavailable for the second study. To rule out off-target binding for the new BRD2 antibody lot, antibody specificity was

confirmed by ChIP sequencing in BRD2 knockout controls (Marit Vermunt, personal communication [data not shown]). These data indicate that global differences in BET protein binding at GATA1 or CTCF do not likely mediate the differential activities of the YFP-tagged BET proteins.

To focus on differentiation-associated binding events, we next performed our chromatin occupancy analysis on peaks that are gained or lost peaks upon GATA1-ER activation. We hypothesized that changes in YFP-BRD2 binding would be mimicked by YFP-tagged BRD3 and BRD4-2, which have BRD2-like activity, but not by BRD4S or BRD2-4S. We thus centered our analysis on YFP-BRD2 peaks. We identified 4,218 YFP-BRD2 binding sites that are lost and 3,080 that are gained upon GATA1-ER activation and measured the average levels of the YFP-BET proteins across these sites. Before GATA1-ER activation, YFP-tagged BRD3 and BRD2-4S levels are similar to those of BRD2 at both lost and gained sites, whereas levels of BRD4S and BRD4-2 are slightly lower, indicating that BRD3 largely mimics BRD2 and that differences in BRD2 and BRD4S levels are determined by the bromodomain-containing halves of these proteins (Figure 4.1.C). Though these differences are maintained upon GATA1-ER activation, all of the YFP-BET proteins follow the same pattern as YFP-BRD2 at sites that are gained and lost. Together, these binding profiles suggest that though there may be subtle binding differences determined by the bromodomain-containing halves of these proteins, extrinsic factors that affect BET protein recruitment – likely histone acetylation – appear to be affecting all of the YFP-BET proteins equally.

However, careful interpretation requires additional testing to determine if histone acetylation patterns differ across the cell lines.

To better understand the changes in occupancy that occur upon GATA1 induction, we identified gained and lost sites for each YFP-BET protein and classified each region by genomic position (intronic, transcription start/termination site [TSS/TTS], etc). This revealed that GATA1-ER activation results in an overall redistribution of each YFP-BET protein towards TSS regions (Figure 4.1.D). For example, TSS regions comprise only ~6% of sites lost by YFP-BRD2 but make up over 30% of the gained sites, whereas the intergenic regions comprise ~50% of lost sites and only ~16% of the gained sites indicating a general shift of YFP-BRD2 away from intergenic regions toward TSS regions. This dynamic is true for the other YFP-BET proteins, though the enrichment at TSS upon GATA1-ER activation is less apparent for YFP-BRD2-4S. To highlight the overall similarity of the chromatin occupancy profiles of each YFP-BET protein, a genome browser view of the *Csf2rb* locus is presented as Figure 4.1.E. The adjacent *Csf2rb* and *Csf2rb2* genes were chosen as BRD2-dependent genes in the mRNA expression studies from Chapter 3. The genome browser view shows enrichment of all five YFP-BET protein at the intergenic *Csf2rb* region upon GATA1-ER activation, with levels of YFP-BET fairly similar and constant at the adjacent loci. In summary, our global analysis of the genomic distributions of the YFP-BET proteins reveals markedly similar binding at both stable and differentiation-associated loci.

Identification of genomic sites differentially bound by YFP-BET proteins

Because our previous analyses did not identify clear differences in YFP-BET binding patterns, we next looked specifically for regions that were occupied by the BRD2-like BET proteins YFP-BRD2/3/4-2 but not by YFP-BRD4S/2-4S. Out of 23,717 consensus regions prior to GATA1-ER activation, there were only 641 sites that were preferentially bound by YFP-BRD2/3/4-2 (FDR<0.05) and 194 sites that were preferentially bound by YFP-BRD4S/2-4S (FDR<0.05) (Figure 4.2.A, left). Likewise, out of 19,921 consensus regions after GATA1-ER activation, 1,207 were preferentially bound by YFP-BRD2/3/4-2 and only 6 were preferentially bound by YFP-BRD4S/2-4S (Figure 4.2.A, right). These numbers indicate that the vast majority of YFP-BET binding sites are shared by all five YFP-BET proteins, which is consistent with our previous analyses. Differentially bound loci exemplifying each condition, respectively, include the *Dock2*, *Csmd3* and *Txlnb* loci (Figure 4.2.B). The specificity of the differentially binding of YFP-BET proteins at these sites in contrast to most other regions is highlighted by the relatively similar levels of YFP-BET at the *Spdl1* TSS adjacent to the differentially bound *Dock2* TSS. Given that BET proteins are known to facilitate transcription factor recruitment, we hypothesized that the differential binding of YFP-BET proteins to this small percentage of regions may be due to the presence of transcription factors unique to each subset of sites. We therefore assessed these regions for transcription factor binding motifs. Interestingly, prior to GATA1-ER activation the YFP-BRD2/3/4-2-specific peaks are enriched most predominantly for the ETS factor motif whereas the YFP-BRD4S/2-4S-specific peaks are enriched only for

the GATA factor motif (Figure 4.2.C, left and middle). These findings are consistent with the role of ETS factors Fli1 and GABPα and GATA factors GATA1 and GATA2 in erythroid cell development and function[163]. Thus, individual BET proteins may associate with these factors separately at some regulatory regions and these interactions may be driven at least in part by the C-terminal halves of BET proteins. After GATA1-ER activation, YFP-BRD2/3/4-2-specific peaks are enriched for several GC-rich motifs (Figure 4.2.C, right). Of note, only the top five of these motifs are shown, though 25 of these GC-rich sequences meet statistical significance (not shown). Each of these GC-rich motifs comprises only a small number of sites but collectively account for a more substantial proportion of the YFP-BRD2/3/4-2-specific peaks. No consensus motifs were enriched in the YFP-BRD4S/2-4S-specific peaks after GATA1-ER activation, perhaps due to the small number of peaks detected. In summary, only a small number of YFP-BET peaks are bound preferentially by either the BRD2-like YFP-BRD2/3/4-2 proteins or the YFP-BRD4S/2-4S proteins. Of this small fraction, the BRD2-like proteins may be binding preferentially at ETS-factor binding sites whereas the YFP-BET proteins lacking BRD2-like activity may be binding at GATA1/2 binding sites prior to differentiation.

The BRD2 and BRD3 CC domains are helical modules that do not dimerize

Our comparative YFP-BET chromatin occupancy data is consistent with many other studies that demonstrate strongly overlapping binding profiles of endogenous BET proteins[55,63,96,102,114,117,118,129,160]. We therefore hypothesized that the ETCC domain, which we previously showed to control BRD2/3-like activity,

may instead facilitate novel protein-protein interactions. We first characterized the structural features of the CC domain to determine whether it possesses helical structure or facilitates homo- or heterodimerization properties as seen with other coiled coils, including motif B[35]. To assess the structure of human BRD2-CC (715-757) and human BRD3-CC (641-688), we expressed and purified these regions and recorded far-UV circular dichroism (CD) spectra of each protein (Figure 4.3.A). The spectra show minima at 222 nm, indicative of the presence of some helical structure. Raussens' method estimates that each of the two peptides contains ~33% helical structure (http://perry.freeshell.org/raussens.html). Because coiled coils often form dimers[159], we assessed whether either the BRD2-CC or BRD3-CC domains can form a homomeric coiled coil using size-exclusion chromatography coupled to multiangle laser light scattering (MALLS). As shown in Figure 4.3.A, both peptides run with molecular weights that are close to the monomeric weights (5.3 kDa and 7.3 kDa, respectively), indicating that they do not significantly self-associate in solution at a concentration of ~100 μM (accounting for dilution on the column). We also assessed the ability of the peptides to associate with each other using the same approach, but again no significant self-association was observed. We conclude that if these sequences form coiled coils, then they must do this by partnering with other proteins.

The ETCC region binds to the PAF and CK2 complexes

To identify partner proteins, we generated GST-fusion proteins and exposed them to nuclear extracts from MEL cells (see Chapter 2: Materials and Methods; Figure 4.4.A). We included the mB region, which is also a putative coiled

coil. We also included two different sets of mutations in the CC predicted to abrogate coiled structure (CCp and CCd). Given that the ET domain is a known protein interaction domain and that the adjacent CC domain functions cooperatively with the ET domain in the domain-mapping studies, we included GST fusion proteins comprised of the ET domain alone or attached to the adjacent CC (its native context) for comparison. Each sample was compared to a GST-only control to determine signal above background. Mass spectrometry (see "Chapter 2: Materials and Methods") revealed enrichment of a total of 392 proteins (log2ratio > 1 over GST alone at FDR < 0.0001, Table 4.1). Several known BET binding partners were among these, validating our approach. For example, ATAD5, CHD8, SMARCA4/BRG1, WHSC1/NSD2, WHSC1L1/NSD3, and CDK9)[64–66,71,76] were enriched significantly by both the ET and ETCC domains (Figure 4.4.B). Likewise, NIPBL, which has recently been shown to interact with the ET domain of BRD4[73,74], was also enriched in these samples. Two other previously reported ET-mediated interactions, CHD4[64,71] and CCNT1[76], did not meet the thresholds set in our analysis but were nonetheless enriched in the ET and ETCC samples. Likewise, mB interaction partner LYAR[83,84] was elevated specifically in the mB sample. The 392 enriched proteins were distributed in the samples as follows: ET (165), ETCC (179), mB (232), CC (0), CCp (2), and CCd (8). These data suggest that the CC does not engage in protein interactions on its own (Table 4.1, Figure 4.4.B).

To identify the proteins that associate most strongly with the GST-fusion proteins, we increased the threshold to log2ratio > 1.5. This identified interactions that had not been previously reported. For example, the ET domain associated

most strongly with AHNAK, AHNAK2, CBX4, ZMYND8 and ZNF592. Interestingly, when combined with the CC domain, ETCC associated with several additional proteins, including all five members of the RNA polymerase II-associated factor (PAF) complex (PAF1, LEO1, CTR9, CDC73, WDR61) as well as the PAF-associated protein IWS1 (Figure 4.4.C, 4.4.D), which are involved in transcriptional elongation[130,164,165]. The ETCC domain also enriched all three members (CSNK2A1, CSNK2A2 and CSNK2B) of casein kinase II (CK2), which like PAF, has been shown to occupy active chromatin[166], associate with elongation factors[167], and contribute to POL2 regulation[168]. The CK2 complex also interacted with mB (Figure 4.4.B). Of note, CK2 has been shown to phosphorylate BRD4 at conserved region that partially overlaps with the mB sequence[85], potentially accounting for its association with BRD2 mB in this assay.

Western blots with antibodies against components of the PAF and CK2 complexes confirmed association with GST-ETCC but less so or not at all with GST-ET or GST-CC (Figure 4.4.E). The ETCC-mediated interactions with PAF and CK2 were also observed when using the ETCC regions from BRD3 and BRDT, but not their ET domains alone (Figure 4.4.F). These results suggest that the BRD2 and BRD3 CC domains augment the ability of the ET domains to engage in specific protein contacts.

Binding patterns of PAF and CK2 on chromatin

The ETCC domain does not appear to alter BET protein binding but may facilitate PAF and CK2 recruitment to chromatin. To assess the feasibility of this mechanism, we measured the genome-wide distributions of the PAF complex

member LEO1 and the CK2 member CSNK2A1 in parental G1E-ER4 cells using ChIP sequencing and compared them to the distributions of YFP-BRD2. We limited our comparison of PAF and CK2 occupancy to that of YFP-BRD2 as the other YFP-BET proteins strongly resemble YFP-BRD2. Technical replicates for CSNK2A1 were highly correlated (Figure 4.5.A). The replicates for LEO1 were also correlated but to a lesser degree, likely due to the use of different chromatin fragmentation methods (see "Chapter 2: Materials and Methods" for details). Thus, in the following analysis, the CSNK2A1 signal from each replicate was merged, whereas for LEO1 each replicate was analyzed separately.

We first assessed the overall binding of CSNK2A1 and LEO1 to various genomic positions to see if either resembled the overall distribution of YFP-BRD2. The distribution of CSNK2A1 more closely resembled that of YFP-BRD2 with >20% of peaks mapping to TSS regions, whereas both LEO1 samples had lower enrichment at TSS regions and greater enrichment at TTS and 3' UTR regions (Figure 4.5.B). The distribution of LEO1 is consistent with the chromatin distributions of other PAF members[130] and with PAF involvement in the elongating POL2 complex as it travels through genes[164]. To visualize these differences in more detail, we generated heatmaps and meta-profiles for each factor across genes. This reveals an accumulation of YFP-BRD2 and CSNK2A1 proximal to the TSS with some extension into the gene bodies (Figure 4.5.C). This distribution of CSNK2A1 is consistent with that of a previous study implicating CK2 directly in transcriptional elongation[168]. In contrast, LEO1 is most notably found in gene bodies toward the end of genes and in the 3' UTR region downstream of the TTS.

Of note, the different profiles of LEO1_r1 and LEO1_r2, particular near the TSS, are consistent with previous reports of LEO1 occupancy using the different fragmentation methods used to prepare these samples. Sonication-based methods (as used for LEO1_r1) yield stronger LEO1 signal near the TSS[130,169], whereas enzyme-based methods (using endo/exo-nucleases, as used for LEO1_r2) yield a relative absence of LEO1 at the TSS[170]. This may be due to digestion of nucleosome-depleted chromatin at the TSS using enzymatic fragmentation. Though the distributions may differ, the cumulative intensity of each factor (both LEO1 replicates, CSNK2A1, and YFP-BRD2) over these regions is positively correlated suggesting co-involvement in gene transcription. The similar distributions of CSNK2A1 and YFP-BRD2 on chromatin support the notion of a stable interaction in vivo, though it can't be ruled out that these factors are recruited independently to genes. The different patterns of YFP-BRD2 and LEO1 suggest that any bona fide interaction may occur transiently or perhaps before both factors assemble on chromatin.

Because CK2 colocalizes with BET proteins on chromatin and has also been shown to phosphorylate PAF, we hypothesized that phosphorylation may at least partially regulate the stability of the ETCC domain bound to PAF. We therefore repeated GST-ETCC pulldown assays using nuclear extracts generated from parental G1E-ER4 cells in the presence of phosphatase inhibitors, which preserve the overall phosphorylated state of the protein lysate. Strikingly, addition of phosphatase inhibitors dramatically hindered the interaction between the ETCC domain and all five members of the PAF complex (Figure 4.5.D). This finding

suggests that CK2, or perhaps one of the other promoter-proximal kinase, may regulate the affinity of the BRD2/3-PAF interaction on chromatin via phosphorylation. This mechanism may account for the reduced TSS-proximal localization of LEO1 on chromatin. Based on our findings, we propose a model by which BRD2/3 contributes to transcriptional regulation distinctly from BRD4 (Figure 4.5.E). We speculate that BRD2/3 functions to recruit PAF to the TSS and that subsequent phosphorylation of the PAF complex by CK2 causes its disassociation from BRD2/3 into the elongating POL2 complex, allowing transcription to proceed.

Discussion

BET proteins are transcriptional regulators that share a conserved domain structure. In the previous chapter, we found that despite this shared structure, BRD2 and BRD3 contribute to transcription independently of BRD4. We traced this unique BRD2/3 function to the ETCC region, which BRD4 lacks. In this chapter, we explored whether the ETCC region may affect BRD2/3 function by influencing where BRD2 and BRD3 bind the genome in comparison to BRD4S. Surprisingly, we found this not to be the case. All three BET proteins bind the genome at GATA1 sites but not CTCF sites and undergo similar changes in occupancy upon erythroid differentiation. Though BRD4S signal is lower at these sites, exchanging the bromodomain-containing segment with that of BRD2, but not the ET/ETCC-containing segment, restores binding to BRD2 levels during erythroid maturation. Together, these data indicate that distinct BET protein functions are not related to differences in genomic localization. Instead, we found that the ETCC domain forms contacts with the PAF and CK2 complexes, both of which have been implicated in

transcription. Though PAF and CK2 colocalize with BET proteins at promoters, the majority of PAF is found over gene bodies and its interaction with the ETCC domain is phosphorylation-dependent, suggesting that it may transiently engage with BRD2/3 in a regulated manner. In summary, we conclude that the ETCC region functionally distinguishes BRD2/3 from BRD4 via the recruitment of distinct partner proteins that regulate transcription.

The lack of colocalization between BRD2 and CTCF in this study is not consistent with our previous observation[131]. Technical differences may have contributed to the discrepancy. For example, our current study used MNase digestion rather than sonication to fragment DNA and used an antibody targeting ectopically expressed YFP-BRD2 instead of endogenous BRD2. These parameters were chosen after several rounds of optimization to yield strong and reproducible ChIP signal. It is conceivable that MNase preferentially degrades CTCF-associated DNA (though unlikely based on preliminary experiments). Likewise, it is possible that the YFP epitope blocks BRD2 binding at CTCF sites. However, more elusive technical differences are likely to blame, as we were unable to re-demonstrate endogenous BRD2 binding at CTCF sites using sonication-based fragmentation and a newer lot number of the initial BRD2 antibody. It is important to note that several other studies have observed some degree of BRD2-CTCF colocalization[122,129]. However, these experiments were performed using the same BRD2 antibody and were published in the same time frame as our previous study, suggesting that this finding may be specific to a particular batch of the BRD2 antibody. Several BRD2 ChIP sequencing studies have been published in the

meantime using various antibodies and experimental approaches[56,102,110,118,171]. Future studies should perform CTCF ChIP sequencing in these models to thoroughly address the robustness of the overlap between BRD2 and CTCF across experimental systems and approaches.

Several studies have reported the distributions of BRD2, BRD3, and BRD4 on chromatin[63,96,102,110,114,117,118,129,160,171]. These studies use antibodies targeting the endogenous BET proteins. Therefore, the signal intensity is dependent on the amount of protein expression and the affinity of the antibody. These biological and technical differences may account for the partially different binding patterns when comparing BRD2, BRD3, and BRD4. Our study overcomes these limitations by using a tightly controlled ectopic system in which each BET protein is expressed at levels comparable to endogenous BRD2 and shares the same epitope tag for immunoprecipitation. Using this approach, we report strikingly similar genomic binding profiles of BRD2, BRD3, and BRD4S. This indicates that when BET proteins are expressed at equal levels and immunoprecipitated with the same antibody, they yield nearly identical chromatin binding patterns. This finding has several implications. First, differences in the affinities of the BET protein bromodomains for various acetylated substrates are unlikely to drive global differences in BET protein localization. Second, differences in BET protein function observed by us and others are not likely due to differential binding. Instead, BET proteins appear to co-congregate at highly acetylated chromatin near promoters and enhancers after which they then exert their unique influences on transcription.

Our study used BRD4S rather than BRD4L for our comparisons to BRD2 and BRD3 because BRD4S structurally resembles these proteins in size. Therefore, our YFP-BRD4S ChIP sequencing data may not be representative of BRD4L, which is the dominant isoform expressed in G1E-ER4 cells and the isoform assayed in virtually all previously published BRD4 ChIP sequencing studies. Because previous studies suggest that BRD2 prefers promoters and BRD4L prefer enhancers, future studies should test whether this holds true when BRD2 and BRD4L are expressed at equal levels and precipitated with the same antibody. If so, our data would suggest that these predilections for genomic regions are driven not by bromodomain interactions but by protein-protein associations mediated by the C-terminal domains. In particular, BRD2 may associate more tightly with PAF and/or CK2 at promoters, whereas BRD4L may associate preferentially with PTEFb at more distal elements[67,172]. Future studies should also address how the ratio of BRD4S to BRD4L is regulated, as BRD4S appears to be the dominant isoform in some cell types[70], and whether this stoichiometry has an effect on transcription.

To investigate how the ETCC domain contributes to BRD2/3-selective function, we identified partner proteins using a proteomic approach. This revealed PAF and CK2 as the dominate ETCC-specific interactions. Because BRD4 does not possess the ET-proximal CC, BRD4 is not predicted to interact with PAF and CK2 by this mechanism. However, CK2 is known to phosphorylate BET proteinsat a conserved serine-rich region that partially overlaps with the mB segment used in this study[85]. Thus, mB-mediated enrichment of CK2 (Figure 4.4.B) likely relates to

this enzyme-substrate binding event and applies to BRD2, BRD3, and BRD4 equally. It is plausible that CK2 phosphorylates serine-rich regions in the ETCC domain of BRD2 and BRD3 in a similar manner, but this is unlikely given that neither the ET nor CC domains alone enrich CK2 and no predicted CK2 substrates span the ET-CC junction (NetPhos 3.1). Thus, whereas CK2 may engage BRD2 and BRD3 via multiple contacts, it likely only binds BRD4 via phosphorylation near the mB region. This is supported by immunoprecipitation data which revealed CK2 is more highly enriched with the CC-containing BRD2/3/T proteins compared to BRD4, and that these interactions are enhanced upon bromodomain inhibition, indicating that they are mediated by C-terminal protein-interaction domains such the ETCC rather than by the bromodomains[31]. Future studies should address which subunit of CK2 directly contacts the ETCC and whether it does so in a manner similar to PAF.

Like CK2, PAF has been found in previous BET protein interaction studies[30,71], but it is notably absent in others[31]. These differences may be due to biological and technical variability between studies and the relatively weak nature of the interaction. For example, immunoprecipitation-based studies are limited by protein abundance, antibody quality, and protein extraction conditions. Even under ideal conditions, such as our GST pulldown assay in which large amounts of bait protein encompassing the interaction surface are used, we find the PAF association to be subject to the presence or absence of phosphatase inhibitors in solution. We also find that in vivo crosslinking (ChIP sequencing) only captures a minority of PAF co-localized with BET proteins on chromatin. Together, these

observations suggest that the BRD2-PAF interaction is transient and highly regulated by phosphorylation and/or other post-translational modifications.

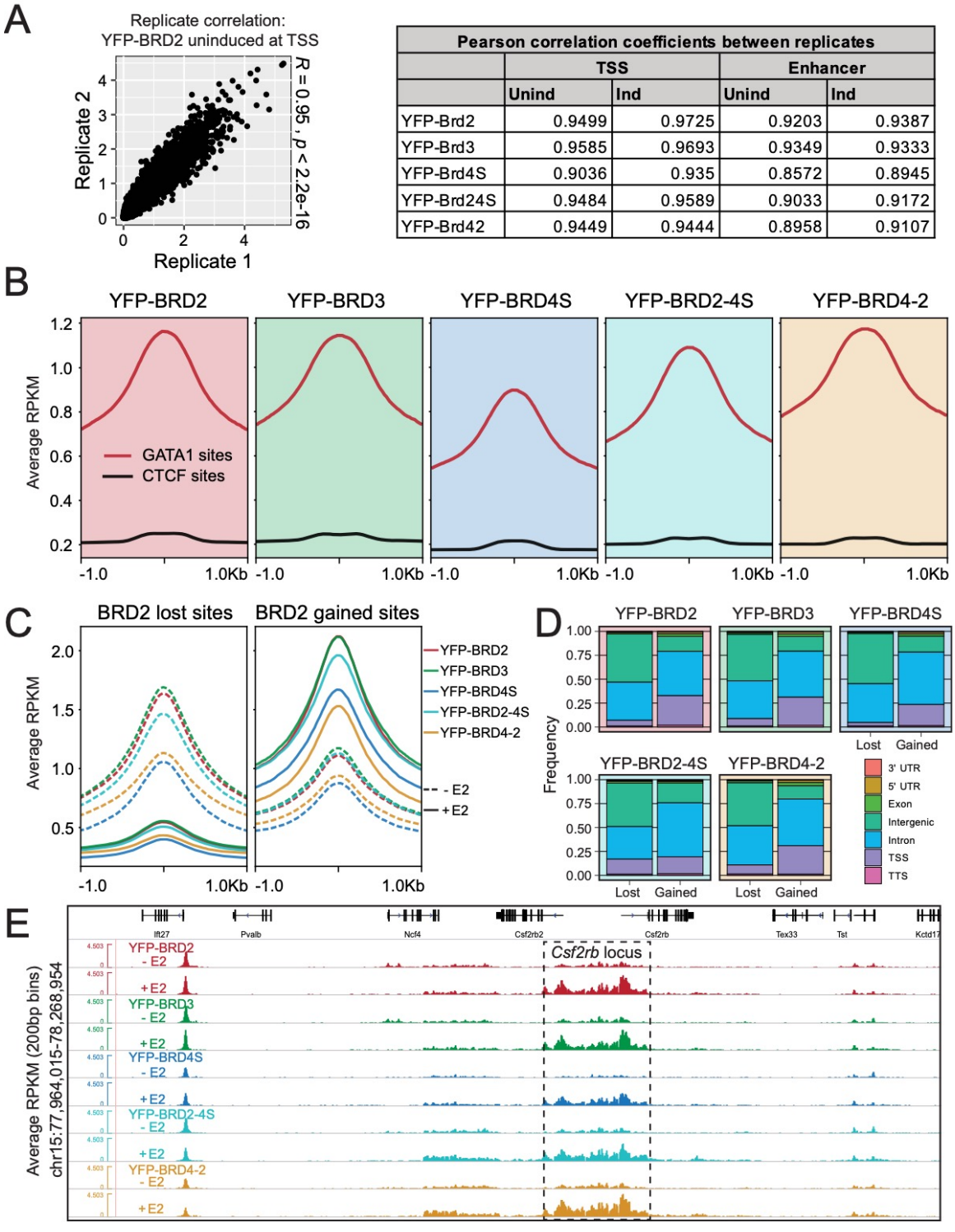
Future studies should confirm that full-length BRD2 and BRD3, but not BRD4, use the ETCC domain to associate with PAF and CK2. Based on the limitations of immunoprecipitation described above, approaches that use highly expressed or purified proteins, such as baculovirus-insect cell expression or in vitro transcription/translation, should be leveraged for this line of investigation. Once the individual partner proteins are identified from these multi-unit complexes, structural studies will be required to fully elucidate the details by which the ETCC binds to PAF and CK2. The role of phosphorylation in regulating the strength of these interaction must be considered through the use of phosphatase inhibitors and exogenous phosphatase. Ultimately, the exact residues that are phosphorylated and the kinases and phosphatases implicated in this regulatory mechanism should be identified and their role in transcription described.

In addition to identifying PAF and CK2 as ETCC-specific interactions, we also identified ZMYND8 as the strongest ET-only interaction. ZMYND8 has been shown to form a highly abundant trimeric complex with ZNF592 and ZNF687 (Z3 complex)[173,174], both of which were also present in our proteomic data (though not as highly enriched). Previous studies have shown ZMYND8 to associate with BET proteins, including the BRD4-NUT fusion protein[30,31,92]. As in our study, this interaction appears to occur via the ET domain[31]. Interestingly, ZMYND8 can function both as a transcriptional activator and as repressor, depending on whether it is in its monomeric or dimeric state[175]. In its dimeric state, ZMYND8 interacts with PTEFb

whereas as a monomer ZMYND8 interacts with the CHD4 component of the nucleosome remodeling and deacetylase (NuRD) complex[175,176], both of which also associate with the ET domain. Further investigation into the role of the ET domain in mediating the various conformations of ZMYND8 and the impact of this on transcription is required. A direct fusion between BRD4 and ZNF592 (a component of the Z3 complex with ZMYND8) mimics BRD4-NUT in the molecular pathogenesis of NUT midline carcinoma[177] indicating that the Z3 complex may be a major mediator of BET protein function. In addition to transcription, ZMYND8 may also coordinate with BET proteins and NuRD to facilitate double-strand break (DSB) repair[127,178,179]. These data support a multifunctional role for BET proteins via the recruitment of diverse chromatin regulators.

The presence or absence of the CC domain may influence the strength of ET-mediated interactions with other protein complexes. For example, ZMYND8/ZNF592, TBLX1, and AHNAK/AHNAK2 are strong ET-only interactions, but their signal is reduced when the CC is included in the ETCC sample. In contrast, PAF and CK2 proteins are more enriched in the ETCC sample compared to the ET-only sample (Figure 4.4.B/C). As mentioned above, ZMYND8 and ZNF592 are components of the Z3 complex. TBL1X is a component of the nuclear receptor co-repressor (NCoR) complex[180]. Several other NCoR members (TBL1X1R1, NCOR1, NCOR2, HDAC2, and HDAC3) are enriched in the ET sample (Table 4.1). Together, these data suggest that whereas PAF and CK2 preferentially associate with BRD2 and BRD3, Z3 and NCoR may preferentially bind to BRD4 which lacks the CC. This differential recruitment of transcriptional

regulators may have functional consequences. For example, preferential recruitment of NCoR, a repressive complex, may explain the growth restriction and reduced gene expression observed when BRD4S is over-expressed in G1E-ER4 cells. The ETCC module may therefore control the major transcriptional mechanisms directed by each BET protein.



Pearson correlation coefficients between replicates				
	TSS		Enhancer	
	Unind	Ind	Unind	Ind
YFP-Brd2	0.9499	0.9725	0.9203	0.9387
YFP-Brd3	0.9585	0.9693	0.9349	0.9333
YFP-Brd4S	0.9036	0.935	0.8572	0.8945
YFP-Brd24S	0.9484	0.9589	0.9033	0.9172
YFP-Brd42	0.9449	0.9444	0.8958	0.9107

Figure 4.1. YFP-BET proteins bind to similar regions of the genome. A) Left: Representative scatter plot comparing chromatin binding signal at transcription start sites (500 bp windows) between two technical replicates with the Pearson

correlation coefficient indicated. Right: Summary table of all of the Pearson correlation coefficients at transcription start sites and annotated enhancers between technical replicates. B) Average YFP-BET chromatin binding signal across 2kb regions spanning all CTCF and GATA1 binding sites[113,131]. C) YFP-BET binding at 2kb regions spanning YFP-BRD2 binding sites that are gained or lost upon GATA1-ER activation. D) Percent/frequency of gained or lost binding sites upon GATA1-ER activation for each YFP-BET protein categorized based on proximity to known genes. E) Genome browser view of the *Csf2rb* locus, representing a region all YFP-BET proteins are recruited upon GATA1-ER activation despite some (YFP-BRD2/3/4-2) having more activity than others in the expression of the *Csf2rb* and *Csf2rb2* genes.

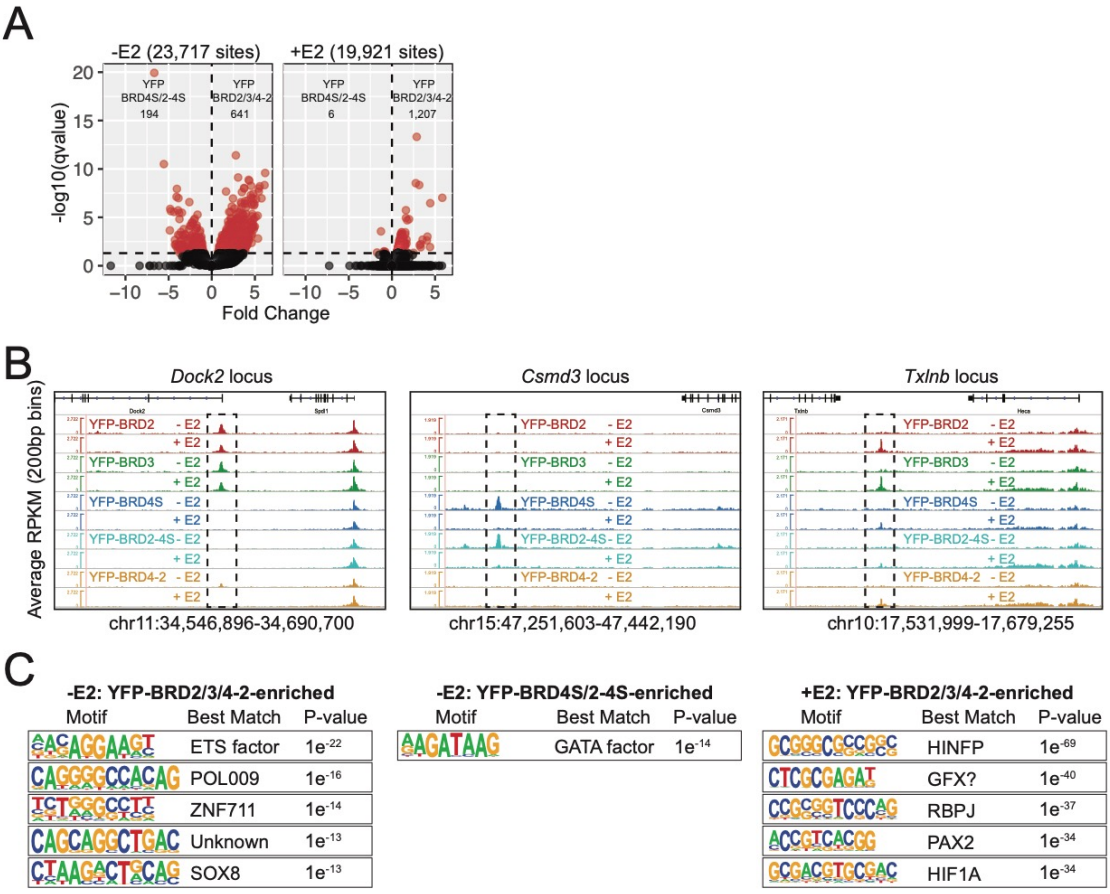


Figure 4.2. Identification of differential YFP-BET binding sites. A) Volcano plots depicting the mean YFP-BET binding signal for YFP-BRD2/3/4-2 in comparison to YFP-BRD4S/2-4S before or after 24 hours of GATA1-ER activation (+/- E2). Each dot represents a differentially bound peak. Red dots are differentially bound peaks at an FDR ≤ 0.05. The number of such peaks is indicated. B) Genome browser views of representative differentially bound peaks. The peaks of interest are in a dashed box. C) Top 5 de novo motif analysis results with motif logos, best matched transcription factors and P-values indicated of the indicated sets of differentially bound peaks in comparison to all identified peaks.

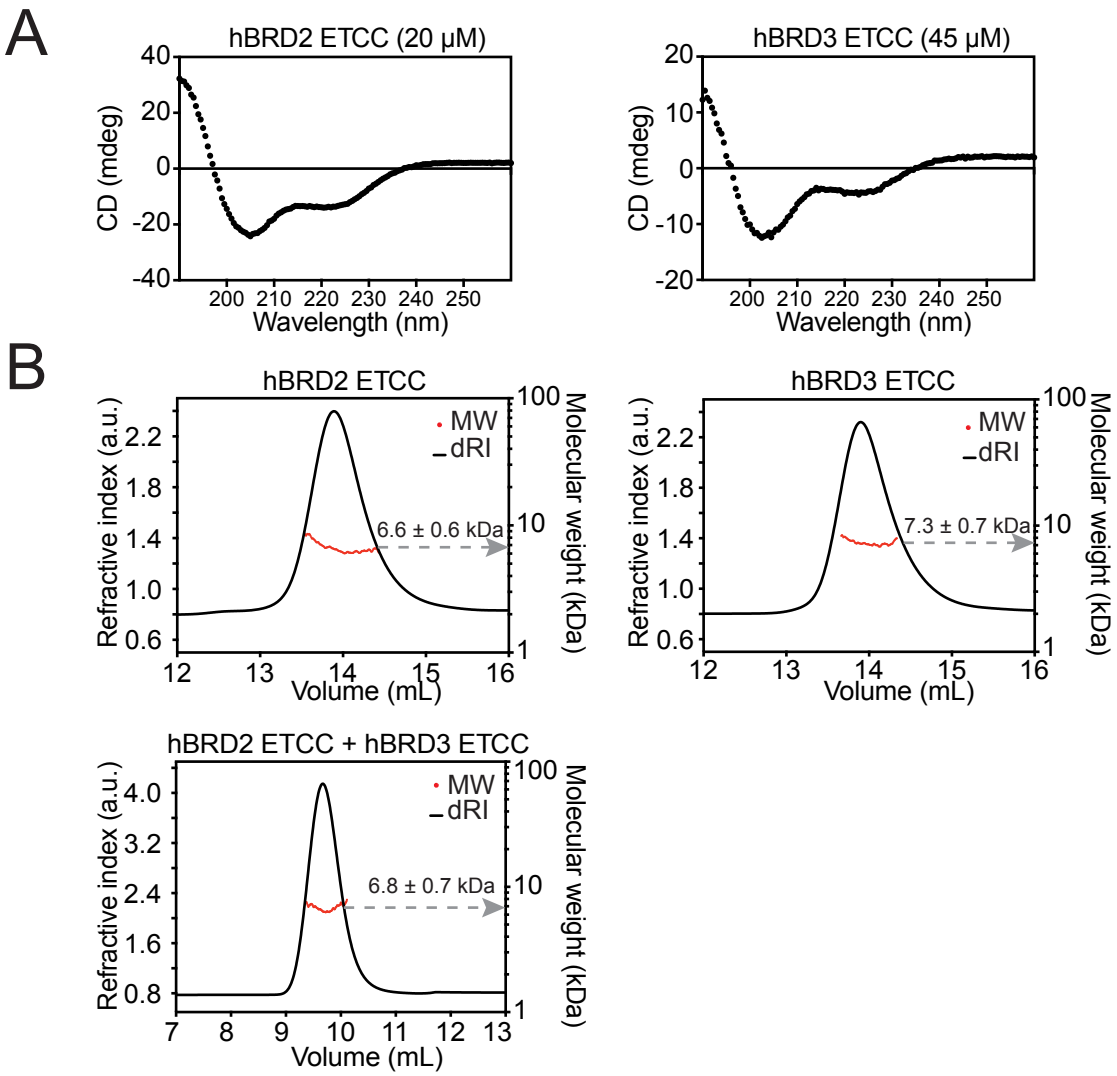


Figure 4.3. The CC has helical structure but does not homo- or hetero-dimerize in solution. A) Far-UV circular dichroism spectra of hBRD2-CC (20 μM) and hBRD3 (45 μM). Spectra were obtained in 10 mM Tris pH 7.2, 100 mM NaCl and 1 mM TCEP at 4 ºC. B) SEC-MALLS data showing that hBRD2 and hBRD3 CC regions are monomers in solution. hBRD2 CC, hBRD3 CC or an equimolar mixture of the two proteins (total protein concentration = 700 μM in each case) was injected onto a Superdex 75 10/300 GL (for hBRD2 CC and h BRD3 CC) or Superdex peptide 10/300 GL (for the hBRD2+hBRD3 CC mixture). The calculated molecular weight is shown together with an estimate of the standard deviation.

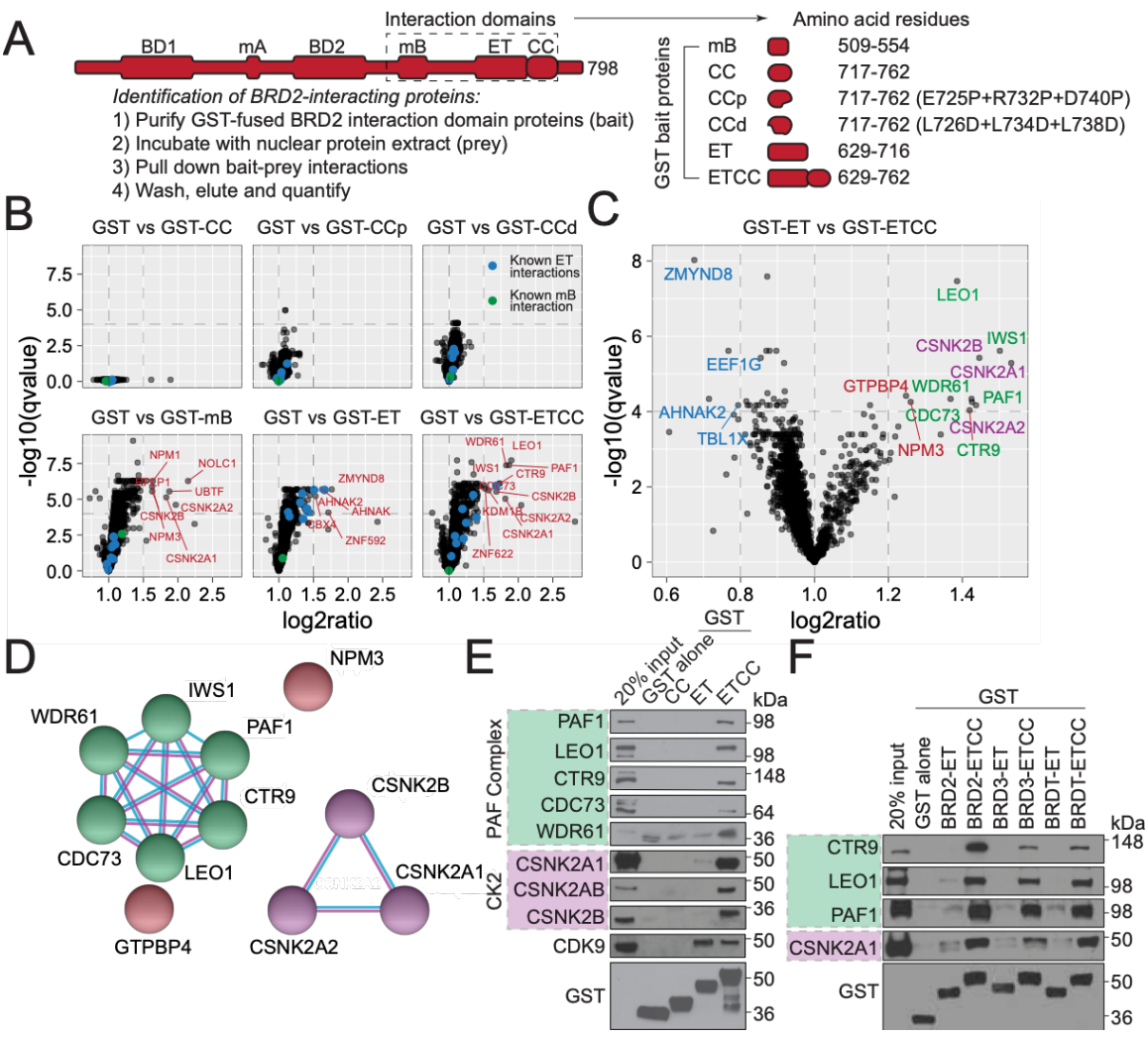


Figure 4.4. The combined ETCC module binds to the RNA polymerase II-associated factor (PAF) and casein kinase 2 (CK2) complexes. A) Experimental schematic of GST pulldown assay. B) Volcano plots depicting enrichment of bait proteins in each sample in comparison to GST alone controls. Prey proteins that exceed a log2ratio of 1.5 and -log10Qvalue of 4 (dashed lines) are labeled in red. Blue and green dots represent proteins known to interact with the ET domain (blue) or mB domain (green) as described in the text. C) Volcano plot comparing prey proteins enriched by the ET domain alone or the combined ETCC. Prey proteins exceeding log2ratio of 1.2 and -log10Qvalue of 4 are labeled in blue if higher in the ET sample and red if higher in the ETCC sample. Members of the PAF complex are colored green and members of CK2 are colored purple. D) Protein interactions (from STRING) of ETCC-high proteins. Purple lines indicate experimental determined protein interactions. Blue lines indicate interactions from curated databases. E) Western blot validation of PAF and CK2 complex members. F) Western blot validation of PAF and CK2 interactions with the ETCC domains from BRD3 and BRDT.

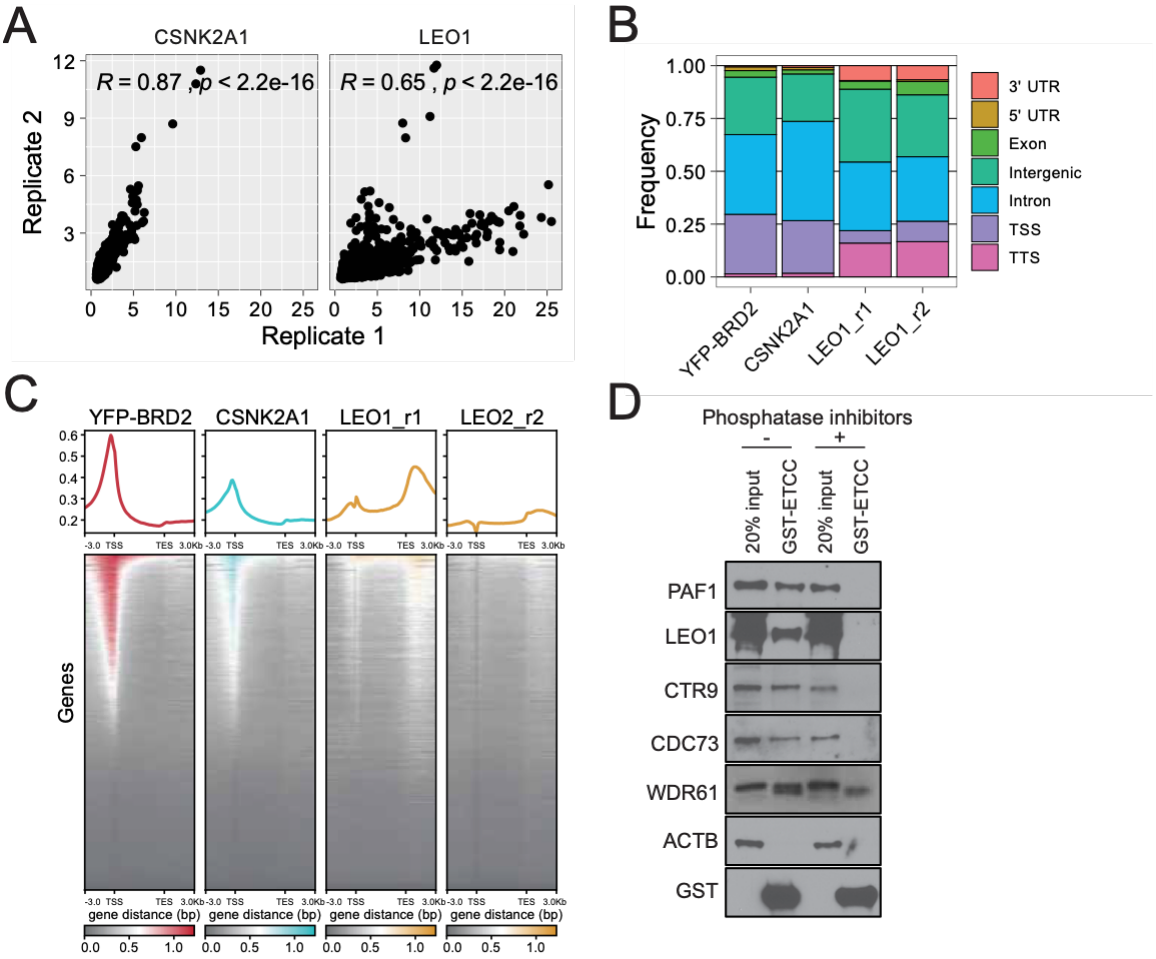

A
CSNK2A1
LEO1
R = 0.87 , p < 2.2e-16
R = 0.65 , p < 2.2e-16
Replicate 2
Replicate 1
B
Frequency
3' UTR
5' UTR
Exon
Intergenic
Intron
TSS
TTS
YFP-BRD2
CSNK2A1
LEO1_r1
LEO1_r2
C
YFP-BRD2
CSNK2A1
LEO1_r1
LEO2_r2
Genes
gene distance (bp)
D
Phosphatase inhibitors
20% input
GST-ETCC
PAF1
LEO1
CTR9
CDC73
WDR61
ACTB
GST
E
Model for BRD2/3-dependent transcription
1) CK2 sits at promoters with BRD2/3/4
2) BRD2/3 transiently recruit PAF
3) CK2 phosphorylates PAF
4) Phosphorylation releases PAF, allowing it to incorporate into elongating POL2
CK2
PAF
BRD2
BRD3
BRD4S
paused POL2
TF
phosphorylation
elongating POL2

Figure 4.5. Comparison of YFP-BRD2, CK2 and PAF localization across the genome. A) Scatter plot comparing technical replicates of CSNK2A1 and LEO1 chromatin occupancy consensus peaks for each factor labeled with Pearson coefficients and p values for the correlation. See text for technical differences between replicates LEO1_r1 and LEO1_r2. B) Percent/frequency of binding sites for YFP-BRD2, CSNK2A1 and each LEO1 replicate categorized based on proximity to known genes. C) Meta-profiles and heatmap matrix illustrating average and individual chromatin binding distributions of YFP-BRD2, CSNK2A1 and LEO1 across all known genes including the flanking 3kb regions. Gene lengths are normalized to 100 bins. D) Western blot detection of PAF members pulled down by GST-ETCC with or without the presence of phosphatase inhibitors. ACTB and GST are included as a negative and positive pulldown controls. E) A speculative model of BRD2/3-based transcription based on the data presented in this dissertation.

Prey proteins detected by each GST-fusion bait protein					
ET			ETCC		
Myb	Rpl35a	Phf8	Tceb1	Rpl6	Grwd1
Prrc2c	Cbx4	Ppp1r10	Xrn1	Rplp1	Mphosph10
Ints3	Hdac3	Med23	Lims1	Rpl5	Smndc1
Lims1	Rpl7	Cwf19l2	Rpl23a	Rpl13	Cggbp1
Hdac2	Srp54	Tbl1xr1	Wdr61	Rpl26	Gltscr2
Polr2a	Rpl13a	Znf592	Rpl19	Rpl27	Rcc2
Smarca4	Hspa5	Mtbp	Zmym4	Rpl37a	Rpl24
Med25	Tln1	Rcc2	Mllt3	Rps7	Rsl1d1
Wdr61	Pik3r1	Ticrr	Mrto4	Rps8	Iws1
Gpatch8	Pura	Lrch3	Tpx2	Rps14	Rbm14
Cdc27	Cct8	Chd9	Ythdc2	Rps23	Utp15
Med14	Hist1h1e	Anapc2	Actb	Rps11	Tceb3
Mllt3	Rpl6	Med19	Rbm34	Rps13	Rbm28
Tpx2	Rplp1	Tbl3	Llph	Rps4x	Gnl3
Raly	Rfx1	Bcor	Aff3	Rpl18a	Kdm1b
Mga	Cit	Ago1	Whsc1	Rps6	Cdc73
Actb	Anapc1	Cdc73	Bre	Tceb2	N4bp2l2
Twistnb	Wdr5	Bclaf1	Mki67	Rpl31	Polr1e
Kif7	Ybx1	Taf10	Rplp0	Rpl32	Pop1
Cep57	Rps27a	Pop1	Nolc1	Rpl8	Pacs1
Mbd3	Ywhah	Paf1	Kif23	Ybx1	Paf1
Anapc7	Erh	Mepce	Rnf169	Actg1	Mepce
Aff3	Smarcc1	Tada3	Ankrd17	Rps17	Ccdc137
Thoc6	Hnrnpul2	Luzp1	Ddx55	Csnk2b	Nop14
Whsc1	Prpf4	Nol6	Rpl15	Erh	Luzp1
Urb1	Larp7	Taf9	Inf2	Bop1	Nol6
Mki67	Tfdp1	Traip	Nop2	Rplp2	Imp4
Gm8730	Chd8	Med28	Rpl29	Sap30bp	Traip
Kif14	Vrk3	Smc6	Rps18	Larp7	Ddx1
Ankrd11	Rfc3	Med1	Taf5	Eif5b	Znf622
Kif23	Gtf3c3	Cdk9	Ddx24	Ect2	Taf5l
Ahnak	Zmynd8	Med24	Zc3h18	Chd8	Arrb2
Ercc2	Mta3	Yaf2	Rpf2	Gpatch4	Ppan
Rnf169	Rfc4	Med30	Tcof1	Mybbp1a	Ddx27
Ddx55	Lsg1	Rnf2	Rpl10	Utp18	Prpf3

Hjurp	Eef1g	Med31	Dnmt1	Rrp12	Cdk9
Gatad2a	Atad5	Mau2	Rpl21	Sf3b2	Gnl2
Med16	Actl6a	Haus5	Dhx15	Lsg1	Mmtag2
Rpl29	Orc2	Ccdc101	Csnk2a2	Rpl4	Gtpbp4
Ahnak2	Thrap3	Ftsj3	Eif6	Nufip2	Npm3
Pcnt	Ttf2	Ipo7	Rpl35a	Npm1	Nop16
Taf5	Kat7	Mllt1	Cbx4	Leo1	Rpl14
Ncor2	Leo1	Aff4	Ncl	Rpl10a	Hnrnpa0
Ddx24	Setd5	Hic2	Rpl7a	Csnk2a1	Med31
Ppp2r1b	Pml	Tbl1x	Dnmt1	Khdrbs1	Rpl11
Brf1	Rbbp4	Exo1	Rpl27a	Kpna1	Rpl34
Rnf219	Rbbp7	Med20	Rpl7	Ctr9	Mau2
Rpf2	Hcfc1	Prpf40a	Rplp0	Top2b	Ebna1bp2
Tcof1	Smarcd1	Racgap1	Hist1h1c	Zcchc2	Pelp1
Opa1	Taf6	Nup160	Rpl13a	FAM120A	Ftsj3
Dnmt1	Zcchc2	Adnp	Rps2	Nop58	Brix1
Hdac1	Arhgap21	Ik	Tln1	Ddx31	Mllt1
Rpl21	Nipbl	Orc1	Rpl3	Bms1	Aff4
Pold2	Whsc1l1	Mbd3	Pabpc1	Rps9	Tmod3
Slk	Tubb2a	Mbd2	Rpl12	Rpl36	Utp3
			Chd1	Rpl17	Rcl1
			Rpl28	Cd3eap	Exo1
			Pura	Ppp1r10	Racgap1
			Hist1h1e	Clspn	Mbd2

Prey proteins detected by each GST-fusion bait protein (continued)					
mB				CCd	CCp
Msh3	Pnn	Rps10	Rbm28	Gpatch8	Mov10
Dhx30	Bysl	Ppp2ca	Sltm	Hnrnpl	Abcf1
Rbpj	Csnk2a2	Csnk2b	Gnl3	Actg1	
Rpl23a	Smarce1	Polr1b	N4bp2l2	Ywhah	
Wdr61	Eif6	Rpl36a	Polr1e	Khdrbs1	
Rpl19	Rpl35a	Erh	Pop1	Cyfip1	
Gemin5	Gtf2h4	Polr1d	Nat10	Rbm14	
Mrto4	Ncl	Rps3a1	Paf1	Ipo7	
Tpx2	Rpl7a	Smn1	Wdr33		
Ubtf	Dnmt1	Rplp2	Sf3b4		
Casc5	Rpl27a	Hnrnpul2	Ccdc137		
Rps16	Rpl7	Cdk7	Pspc1		
Slfn8	Rpsa	Top1	Nop14		
Lrrc47	Hist1h1c	Akap8	Nol6		
Ythdc2	Srp14	Larp7	Znf385a		
Actb	Rpl13a	Eif5b	Hnrnpr		
Twistnb	Rps2	Gpatch4	Traip		
Ddx18	Rpl3	Srsf10	Rbmxl1		
Rbm34	Rpl12	Fam98a	Ddx1		
Gfi1b	Chd1	Mybbp1a	Znf622		
Cbfa2t3	Rpl28	Dazap1	Arrb2		
Cep57	Hist1h1e	Rrp12	Gtf2h2		
Ylpm1	Hist1h1a	Eif2s2	Sf3b3		
Llph	Rpl6	Lsg1	Ddx27		
Rps5	Rplp1	Serbp1	Smc6		
Sf1	Rpl5	Polr2e	Eif3m		
Urb1	Rpl13	Nmnat1	Gnl2		
Msto1	Cit	Ddx47	Mmtag2		
Ankhd1	Ercc3	Rpl4	Gtpbp4		
Farsa	Polr1c	Nufip2	Npm3		
Gm8730	Snrpa1	Npm1	Nop16		
Nolc1	Fam207a	Leo1	Nop10		
Kif23	Rpl26	Rpl10a	Rpl14		
Ercc2	Rpl27	Csnk2a1	Nhp2		

Rnf169	Rpl37a	Fxr1	Rpl11		
Ankrd17	Rps8	Srsf2	Ilf2		
Ddx55	Rps15a	Snrnp70	Skiv2l2		
Fmr1	Rps14	FAM120A	Rps19		
Rpl15	Rps23	Ncapg2	Ndc80		
Gtf2h1	Rps29	Nop58	Rpl34		
Nop2	Rps11	Pdcd11	Nop56		
Dhx9	Rps13	Ddx31	Ebna1bp2		
Rpl29	Rps4x	Rps9	Ftsj3		
Rps18	Rpl18a	Rpl35	Brix1		
Srbd1	Rps6	Rpl36	Aff4		
Cdc45	Rps15	Rps12	Utp3		
Ddx24	Rps24	Rpl17	Ddx21		
Zc3h18	Rps25	Cd3eap	Rcl1		
Rpf2	Tceb2	Znrd1	Ddx20		
Trip12	Rpl30	Safb2	Hic2		
Tcof1	Rpl39	Smndc1	Plec		
Knop1	Rpl31	Noc4l	Uchl5		
Rpl10	Rps3	Gltscr2	Racgap1		
Dnmt1	Rpl32	Rcc2	Ik		
Rpl21	Rpl8	Rpl24	Orc1		
Polr1a	Rps27a	Naa25	Ilf3		
Srsf5	Actg1	Utp15			
Kpna6	Rps17	Smc5			

Table 4.1. GST protein pulldown assay results. Target/bait proteins from nuclear extracts pulled down by each GST-fused prey proteins (GST-ET, GST-ETCC, GST-mB, GST-CCd, GST-CCp are listed. No proteins were detected in the GST-CC sample. Protein enrichment was measured using standard mass spectrometry label-free quantification (See “Chapter 2: Materials and Methods”). Only proteins with enrichments exceeding a log2ratio>1 over GST-alone controls at FDR<0.0001 were included.

CHAPTER 5: CONCLUSIONS, LIMITATIONS, AND FUTURE DIRECTIONS

Chapter summary

BET proteins are transcriptional regulators implicated in multiple developmental and pathological pathways. As such, they have emerged as attractive therapeutic targets for many diseases, but their pharmacological inhibition can impair normal processes. Because the current inhibitors typically target the highly conserved BET bromodomains, many of these drugs cannot discriminate between individual BET proteins. Fundamentally, it remains unclear how individual BET proteins contribute distinctly to transcription and whether targeting single BET protein family members would be of therapeutic value. In this study, we set out to better understand the extent to which BET proteins BRD2, BRD3, and BRD4 differ functionally and to discover the domains and mechanisms by which they do so. We used the G1E-ER4 erythroid cell line, a model system where BET proteins are critical to cell growth and GATA1-mediated differentiation and found a key functional domain that may mechanistically distinguish BRD2 and BRD3 from BRD4. We posit that targeting this region, termed the ETCC, may provide a BRD2/3-specific drug that prevents the recruitment of certain transcriptional regulators. In this chapter, we expound upon this concept, discuss limitations of this work, and propose future studies to examine the mechanistic detail and broader applicability of these findings.

Distinct BET protein functions revealed by gene rescue and domain mapping

Previous studies have shown that depletion of individual BET proteins results in distinct phenotypes (see “Chapter 1, Evidence for distinct BET protein functions”). In support of BET protein family members functioning distinctly, our work in G1E-ER4 cells suggests that BRD2 and BRD4 have independent roles in GATA1-mediated erythroid maturation. BRD3 behaves more similar to BRD2 in this context. We draw this conclusion based on gene rescue studies performed by ectopically expressing BET proteins in BRD2 knockout cells. An important limitation to this work is the lack of reciprocal experiments performed in BRD4-depleted cells. We have been unable to generate BRD4 knockout G1E-ER4 cells, perhaps because BRD4 is required for cell survival. However, decreased BRD4 expression and reduce erythroid maturation using RNA interference (RNAi) can be achieved[113]. Future experiments should be designed to test whether ectopic BRD2 and BRD3 expression can protect cells against defects associated with BRD4 depletion using this method. Based on our hypothesis, we would expect that neither BRD2 nor BRD3 can compensate for BRD4 function.

It will be interesting to note whether ectopic BRD4S or BRD4L in isolation will protect against RNAi-related defects. To the best of our knowledge, BRD4L is the dominantly expressed isoform in this cell type, but BRD4S transcripts are expressed and BRD4S has been implicated in transcription in other cell types. For example, in CD4+ T cells, BRD4S is abundantly expressed and recruits SWI/SNF remodelers[70]. In cancer cells, BRD4S levels vary but correlate with nuclear puncta formation and exert BRD4L-independent effects on gene expression[181]. In some

cancers, BRD4S may oppose BRD4L or function independently to promote tumor growth and metastasis[81,182]. Thus, it is conceivable that both isoforms are necessary for erythroid maturation, which can be teased apart using a combination of isoform-targeted RNAi and ectopic BRD4S/L expression.

Identifying the structural basis for overlapping BRD2 and BRD3 function

In the absence of a distinct structural domain, it remained unclear prior to our study how BRD2 and BRD3 could function separately from BRD4, especially given that the bromodomains and ET domains are well-conserved across the family. Our domain mapping approach revealed that the extended ET domain, containing a coiled coil (ETCC) in BRD2 and BRD3 but not in BRD4S, is the critical region distinguishing these proteins. This finding contributes to a mechanistic model of how BRD2 and BRD3 can function distinctly from BRD4. For example, BRD4-specific activity is generally attributed to the CTM on the extended C-terminal tail of BRD4L. Based on our domain mapping, we can now attribute BRD2/3-specific activity to the ETCC domain. Whereas it is thought that the CTM enables BRD4 to strongly associate with PTEFb, we propose that the ETCC domain enables BRD2 and BRD3 to more strongly associate with PAF and CK2. These distinct functions are likely in addition to shared functions, such as binding acetylated chromatin via the conserved bromodomains and recruiting NSD3[65] and JMJD6[67] by the conserved ET domains. We therefore suggest a model in which individual BET proteins function distinctly but interdependently to coordinate the recruitment of multiple transcriptional regulators to chromatin.

Several additional gene rescue studies are needed to bolster this model. First, additional chimeric BET proteins should be designed to test the ETCC domains from BRD3 and BRDT. Based on their ability to bind PAF and CK2, we hypothesize that the ETCC from these proteins should function similarly to the BRD2 ETCC in their ability to impart activity to BRD4S in gene rescue assays. Chimeric BET proteins possessing the BRD4L tail containing the CTM should also be constructed. We might expect that addition of the BRD4L tail to BRD2 would impart BRD4L-like activity to BRD2. In effect, this construct would resemble BRDT, which has the key structural features of both BRD2 (the ETCC) and BRD4 (the CTM). We did not include BRDT in our gene rescue assays, but it's potential pan-BET functionality in G1E-ER4 cells based on possession of both the ETCC and CTM domains may be a useful comparison in future experiments. BRDT is not expressed in G1E-ER4 cells, but we hypothesize that its ectopic expression might compensate for both BRD2 and BRD4 loss.

Structure-function considerations related to BRDT

As noted above, BRDT possesses both the ETCC domain shared by BRD2 and BRD3 and the CTM domain of BRD4. The function of BRDT in comparison to the other BET proteins warrants further exploration. In the testis, where it is exclusively expressed[19], BRDT is required for proper spermiogenesis[24] likely due to an ATP-independent structural role it plays in large scale chromatin re-organization involving histone compaction and protamine replacement[183–186]. Though the pronounced nuclear reorganization that occurs during spermatid formation highlights the role of BRDT in chromatin structure, other BET proteins

have been also been implicated in chromatin organization. For example, the yeast homologue BDF1 maintains euchromatin by preventing heterochromatin spread at transcriptional boundaries[187], and mammalian BRD4 regulates nucleosome density[79,188]. Moreover, BRD2, BRD3, and BRD4 all exhibit *in vitro* histone chaperone activity that is independent of the CTM (in the case of BRD4) and specific for acetylated histones[40,41]. Thus, though it is tempting to ascribe a unique role for BRDT in this process, the other BET proteins may contribute to structural changes during spermiogenesis and may mediate similar acetylation-dependent nucleosomal dynamics in other cell types.

In addition to its structural function, BRDT also mediates testis-specific gene activation via CTM-recruitment of PTEFb to gene promoters, akin to canonical BRD4-related transcriptional regulation[186]. Thus, BRDT likely contributes to spermiogenesis in two fundamental ways: 1) activation of genes required for meiosis and 2) nuclear compaction and histone removal. Interestingly, other BET proteins are expressed in the testis and contribute to spermiogenesis. In fact, *Brd4* heterozygous null mice have reduced spermiogenesis[23], and BRD4 was shown to co-bind BRDT-bound promoter of active genes in spermatids[189]. Not unexpectedly, BET bromodomain inhibition using small molecule JQ1 results in meiotic arrest and reduced fertility in male mice[190]. The fact that BRDT and BRD4 are both required in the testis suggests that they have undergone some level of functional specification. For example, comparison of BRD4 and BRDT genome occupancy in post-meiotic spermatids revealed that only a small percentage of BRDT was found at gene promoters in comparison to BRD4[189]. And whereas BRD2 and BRDT

localize diffusely throughout the nucleus in the late stages of spermatid development, BRD4 forms a distinct ring-like structure adjacent to the cytoskeletal acrosome in a confined band around the nuclear periphery[189]. Of note, it is unclear if these structures involve full length BRD4 or a smaller noncanonical, possibly testes-specific isoform. This technical caveat may be meaningful given that the canonical BRD4S isoform associates preferentially with the nuclear envelope whereas BRD4L concentrates in the salt-extractable nuclear matrix[82]. Thus, differences in BET protein localization, genome occupancy, and function due to testes-specific regulation of endogenous proteins is unlikely. Instead, such dramatic differences likely reflect intrinsic protein attributes such as the gain/loss of entire domains. For example, the intranuclear compartmentalization of BRD2 and BRDT versus the acrosomal localization of BRD4 might be explained the shared possession of the ETCC by BRD2 and BRDT but not BRD4. Thus, future studies should compare these proteins head-to-head in G1E-ER4 cells using equally expressed and epitope tagged constructs. This may lead to new insights about how the CTD-containing proteins BRD4L and BRDT function distinctly from one another in comparison to BRD2, BRD3, and BRD4S.

In the absence of a robust cell culture system for modeling spermiogenesis, it may be difficult to tease apart the unique or overlapping roles of each BET protein – and map these similarities/differences to certain domains – as we have in the G1E-ER4 system which allows for gene rescue experiments. Nevertheless, spermiogenesis provides a unique biological setting to track differences in endogenous BRD2, BRD3, BRD4, and BRDT genomic distributions and protein

interaction networks through each stage of sperm development. The importance of BET protein function in sperm development is showcased in Drosophila, which lack a testis-specific paralog. In this setting, two proteins critical for spermatocyte development, tBRD-1 which lacks an ET domain and tBRD-2 which lacks a second bromodomain, can dimerize to reconstitute a BRDT-like BET isoform[191]. Future studies examining their individual and combined function can help isolate the critical domains necessary for BRDT's role and may provide a window of insight into general BET protein function.

Mechanisms of PAF recruitment by BET proteins

Validation and further characterization of ETCC-mediated interactions with PAF

We found that the GST-ETCC fusion proteins from BRD2, BRD3, and BRDT were able to precipitate the PAF complex from solution. This reproducibility across conserved but not identical ETCC sequences reassures us that is not a spurious finding. However, several experiments are required to fully validate the functional significance of this interaction. First, point mutations in conserved residues comprising the CC should be explored for their ability to reduce PAF binding without interrupting ET only-interactions. We have preliminary identified a set of proline insertions predicted to abrogate the coil (mutations depicted in Figure 4.4.A as CCp but data not shown) that when placed into ETCC reduce PAF signal without altering CDK9, an ET-only interaction. Interestingly, the leucine to aspartic acid mutations (depicted in Figure 4.4.A as CCd but data not shown) are also predicted to disrupt the coil but do not reduce PAF enrichment. Thus, the effect of these mutations on the structure of the ETCC domain should be verified using

circular dichroism to determine the role of the helical component on these interactions. Second, after isolating point mutations that abrogate the PAF interaction without compromising the ET domain, BRD2 constructs harboring these mutations should be expressed in *Brd2*-null G1E-ER4 cells so that their function in cell growth and GATA1-mediated differentiation can be assessed. Last, PAF occupancy should be measured in these BRD2 ETCC mutant-reconstituted cells to determine the impact of the ETCC mutations on PAF recruitment to BRD2-bound genes.

The GST pulldown experiments used large, micromolar amounts of GST-fused bait proteins to identify partners, in essence flooding the reaction with the interaction surface of interest. One advantage of this strategy is that the domain of interest has already been defined functionally and thus large amounts of it can be used to uncover weaker but still meaningful protein interactions that are perhaps overlooked when full-length BET proteins are used. A disadvantage, however, is that the ETCC domain in this case has been taken out of the context of the properly folded full-length protein, and thus interaction surfaces on the ETCC may be exposed that do not normally occur. Further complicating the matter is that transient interactions may be tightly regulated through post-translational modifications that alter protein conformation to open and close an interaction surface. For example, a phosphorylation-induced intramolecular switch was proposed to toggle BET proteins between two different conformations, one of which exposes the second bromodomain to form contacts with acetylated chromatin[85]. These factors must be considered when interpreting the following

observations: 1) despite robust interaction with the ETCC in our GST pulldown assays, PAF members are inconsistently co-immunoprecipitated with BET proteins[30,31,71], 2) only a fraction of PAF co-localizes with BET proteins on chromatin, and 3) the ETCC-PAF interaction is sensitive to phosphorylation. Future studies using orthogonal approaches and not limited to a single modality[192] are required to identify the exact conditions under which these proteins associate and the implication of these interaction on transcription.

Toward this goal, we recommend against co-immunoprecipitation as the method of choice for further biochemical characterization. In our hands, these assays have failed to yield consistent results despite multiple attempts optimizing the antibody/epitope and buffer conditions. Moreover, immunoprecipitation enriches all three BET proteins regardless of the target BET protein. This is likely due to the tendency of BET proteins to agglomerate with shared protein complexes[30] or to polymerize directly via the motif B[35] rendering it difficult to tease apart which protein interactions are directly mediated by a given BET protein. Immunoprecipitation is also limited by endogenous protein level and specificity of available antibodies. Use of epitope-tagged protein could overcome both limitations but may lead to overexpression artifact.

Instead, we propose using a baculovirus vector and insect cell expression system to purify high yields of individual BET proteins and PAF members (including IWS1). By selectively adding individual components to the reaction mixture, this system can be leveraged to 1) validate that interactions with PAF are strongest with full length BRD2/3/T in comparison to BRD4, 2) confirm that the ETCC domain

is the critical BET protein interaction surface, 3) determine which PAF member directly mediates this interaction, and 4) identify the reciprocal interaction surfaces on PAF. Though the PAF complex can be reconstituted using proteins derived from both bacterial and insect cell expression systems[164,165,193], insect cells retain most post-translational modification processes that may be critical for the interaction with BET proteins.

In vitro transcription/translation-coupled reactions may provide an alternative to the baculovirus system described above. These systems are generally conducted in reticulocyte lysates and may mimic, at least to some extent, the conditions of erythropoiesis. Moreover, these lysates should be devoid of nuclear components and lack endogenous PAF proteins, an advantage over insect cells where the insect PAF homologs may participate in the interactions. This system was used to successfully reconstitute the NuRD complex[194].

Given that coiled coil domains have a tendency to form dimers[35,175], we might target our search for the ETCC-interacting component of PAF to those protein segments that possess putative coiled coil regions. Alternatively, it is also conceivable that all of the ETCC interactions with PAF are in fact indirectly mediated by mutual associations with CK2, though this is unlikely given that PAF and CK2 associate via FACT[167,195], whose components (SUPT16H and SSRP1) are not found in our list of ETCC-enriched proteins.

Functional significance of the BET-PAF interaction

The PAF complex is a well-studied transcriptional regulator shown to affect POL2 pause-release and elongation. During the transition from transcriptional

pausing to elongation, PAF replaces NELF at a mutually exclusive region of POL2, and thus PAF is a core component of the elongating but not paused POL2 structure[164,196]. Within the structure of elongating POL2, PAF induces a conformational change that unwinds upstream DNA, permitting POL2 forward passage[164]. Consistent with its role in elongation, PAF facilitates transcriptional elongation through chromatin templates in vitro[165] and is found enriched over gene bodies on chromatin[130,169]. PAF1 depletion results in decreased elongation[130,197], though some studies show the opposite effect (depletion of PAF increases elongation)[198–200] so the exact role of PAF in transcription remains controversial. PAF has also been implicated in post-transcriptional RNA processing[169,201,202], though studies in yeast show that almost all the effects of PAF depletion on transcripts levels are related to primary transcription suggesting that this is the ancestral function of PAF[203].

Mechanistically, the intrinsic elongation activity of PAF in vitro depends on histone acetylation[165], indicating that PAF may cooperate with factors that possess acetylation-dependent histone chaperone activity. Intriguingly, this type of activity has been observed for BET proteins[40,41]. As the elongating POL2 complex approaches a nucleosome comprised of acetylated histones, BET proteins may be influenced by PAF via the ETCC on BRD2 and BRD3 to help disassemble and reassemble the nucleosome as POL2 passes through. Future studies will need to test whether addition of both PAF and BET proteins to these in vitro transcriptional assays yields synergistic effects. It is important to note that unlike serine-2-phosphorylated POL2 (which demarks the elongating fraction of POL2) and the

associated PAF proteins[130], BET proteins are not found highly enriched in gene bodies on chromatin, indicating that they likely do not travel with POL2 through genes as part of this complex. Instead, only a small fraction of the total BET proteins, which are mostly found near the TSS in promoters or at distal regulatory enhancer elements[55,129], may be bound to nucleosomes within gene bodies. Nevertheless, the density of BRD4 in the gene bodies correlates better with gene expression than does the density of BRD4 at the promoter, indicating that this pool of BRD4 in the gene bodies, albeit the minority of BRD4, may be more directly involved in transcription[41]. A comprehensive correlation of BRD2, BRD3, and BRD4 levels in various genomic regions with gene expression in additional cell types is needed to better establish this paradigm.

In addition to cooperating with the PAF complex to facilitate POL2 processivity through nucleosomes, BET proteins may also recruit PAF to gene promoters. This is where the majority of BET proteins reside on chromatin. A similar mechanism has been proposed for PTEFb recruitment by BRD4 via its CTM-containing tail. But unlike PTEFb, which contributes to elongation but does not incorporate into the elongating POL2 complex, PAF binds to the elongating POL2 complex and leaves the promoter as it travels with POL2. In fact, phosphorylation of POL2 by PTEFb in the promoter dismisses NELF from the paused POL2 complex which allows PAF to takes its place in the elongating POL2 complex[164,196]. These differences in PTEFb and PAF functions are reflected in their genomic distributions. PTEFb sits at the promoters colocalized with unphosphorylated POL2 (marking the paused fraction of POL2), whereas PAF is

split between a population colocalized with PTEFb and paused POL2 at the promoter, a fraction traveling with serine-2-phosphorylated POL2 over gene bodies (marking the elongating fraction of POL2), and a third pool enriched over the TES[130,169]. Given that PTEFb is required for the proper loading of PAF onto POL2, it may be difficult to disentangle the functions of BRD4 (which presumably recruits PTEFb) and BRD2/3 (which we propose to recruit PAF) based on gene expression signatures after depletion of a target BET protein. Instead, changes in the PTEFb and PAF chromatin occupancy patterns described above may be more informative. For example, we would expect that depletion of BRD2 may reduce PAF signal over gene bodies without altering PTEFb, whereas depletion of BRD4 may reduce both PTEFb signal at the promoter and PAF signal throughout gene bodies.

It has also been reported that PTEFb exists in multiple discrete complexes[204], some of which are recruited by BRD4 and others of which are directly recruited by PAF via a mutual interaction with the elongation factor ENL[205]. Thus, rather than BRD2/3 and BRD4 functioning coordinately to promote POL2 pause-release, it is also possible that BRD2/3 is responsible for PAF-dependent PTEFb recruitment whereas BRD4 is responsible for PAF-independent PTEFb recruitment. We would therefore predict BRD2/3-dependent genes to have stronger levels of ENL at promoters whereas BRD4-dependent genes may not, but this remains to be tested. Ultimately, these pathways may be more intertwined then these models suggest. In fact, mutations in either ENL[206] or CTR9[207] (a PAF component) predispose patients to Wilms tumors of the kidney. Given the rarity of

this cancer, this coincidence suggests very similar molecular pathways involving both proteins.

Phenotypic studies exploring the functions of PAF and PTEFb argue in favor of more distinct functions. For example, PTEFb- or PAF-targeted depletion reveals their antagonistic roles in maintaining multipotent neural crest progenitors[208] as well as the pluripotent state of embryonic stem cells[209,210]. Similar opposition between PAF and PTEFb has also been reported in zebrafish for genes controlling oligodendrocyte differentiation[211] and perhaps most relevant to our study, for erythropoiesis[198]. These studies suggest a model in which PAF represses PTEFb-mediated pause-release (preventing expressing of differentiation-related genes) as proposed by Chen et al[199,200]. PAF may therefore have two distinct and seemingly paradoxical functions in transcription. It may poise POL2 in its paused state, and then upon PTEFb phosphorylation of paused POL2, incorporate into the elongating POL2 complex to help facilitate POL2 progression through chromatin. This then allows another paused PAF/POL2 unit to take its place. These dichotomous functions, further complicated by potential auto-feedback loops regulating PAF and PTEFb recruitment and POL2 processivity, may explain the discrepant effects of PAF-depletion on elongation observed by different groups. In summary, the molecular mechanisms underpinning PAF and PTEFb coordination versus opposition are not clear but BET proteins may contribute to these regulatory pathways via BRD2/3 recruitment of PAF and BRD4 recruitment of PTEFb. This may explain why targeted depletion of single BET proteins can elicit unique gene expression changes and cellular phenotypes. Further biochemical characterization

of the BRD2/3-PAF interaction may allow for the development of molecular tools to purposefully and specifically disrupt this pathway to better identify their unique role in transcription.

In addition to its roles regulating POL2 pause-release and elongation, PAF has also been strongly implicated in various histone modifications, particularly ubiquitylation of H2B (H2Bub) and subsequent H2Bub-dependent methylation of H3K4 and H3K79[202]. Mechanistically, H2Bub deposition may occur directly via PAF-mediated recruitment of ubiquitin ligases or indirectly through transcription-mediated recruitment of the ubiquitylation machinery[202]. Regardless, PAF-regulated ubiquitylation H2B presents an intriguing parallel with BRD2 function. Ubiquitylation of H2A.Z is reported to directly antagonize BRD2 binding to acetylated H2A.Z at gene promoters[59]. Given that acetylated H2BK34 and H2BK120, the same residues targeted by H2B ubiquitylation, can bind some of the BET bromodomains[44], it is possible that a similar antagonism between acetylation and ubiquitylation on BRD2 recruitment occurs on H2B, though further studies are required to clarify this. H2Bub may also directly regulate PTEFb recruitment[212], further intertwining the functions of PAF, PTEFb, and individual BET proteins.

Mechanisms of CK2 recruitment by BET proteins

Validation and further characterization of ETCC-mediated interactions with CK2

Interactions between BET proteins and CK2 are well supported by other studies. Dawson *et al.* found that CK2 (included as part of the NOLC1 complex with UBTF, NOLC and NHP2) co-immunoprecipitated with BRD2, BRD3, and BRD4[30]. Wai *et al.* likewise identified CK2 among many other transcription-related

complexes associated with BRD3 (BRD2 and BRD4 were not included in the study)[71]. Intriguingly, Lambert *et al.* identified CK2 as a major interaction partner of BRD2, BRD3, and BRDT but not BRD4[31]. This is consistent with BRD2/3/T but not BRD4 containing the ETCC segment which we propose to at least in part mediate the CK2 interaction. Moreover, inhibition of the BET bromodomains actually enhanced the CK2 interactions with BRD2/3/T, supporting our hypothesis that these interactions are bromodomain-independent. Importantly, CK2 is a serine-threonine kinase and CK2 has been previously shown to phosphorylate mammalian BET proteins as well as the yeast paralogs BDF1/2 at two conserved serine-rich regions[85,213,214]. BRD2 is also one of about 60 candidate CK2-target proteins identified in a large-scale meta-analysis of interlaboratory data[215]. Conservation of the residues phosphorylated by CK2 in BET proteins strongly suggests that these post-translational modifications are of functional significance. One of the conserved CK2 substrates overlaps partly with the motif B (mB) domain used in our GST-pulldown assay. This enzyme-substrate binding event appears to have been captured by the GST-mB sample in our GST pulldown experiment (Figure 4.4.B). The other conserved CK2 substrate occurs immediately 3' to the ET domain in BRD4 and the ETCC domain in BRD2/3/T and in the last 26 residues of yeast BDF1, beyond the coiled coil region. Deletion of either the mB-proximal or ET/ETCC-proximal phosphorylated region impairs the ability of BDF1 to rescue the viability of *bdf1/bdf2*-null strains but does not impair its ability to rescue temperature sensitivity of *bdf1*-null strains, indicating that though both regions are required for full BDF1 activity, some redundancy exists between BDF1 and BDF2

in regard to phosphorylation-specific function[214]. In BRD4, the mB-proximal CK2-substrate functions as phospho-switch regulating bromodomain-dependent chromatin contacts. The purpose of the ET-proximal substrate remains untested in mammalian BET proteins[85,213].

Though putative CK2 phosphorylation sites also occur in the ETCC, we do not believe the ETCC domain itself to be a true CK2 substrate. For example, neither the ET nor the CC domains alone precipitate CK2, indicating that the mechanism of CK2 interaction with the adjoined ETCC is likely to be more related to structural characteristics of the surface created by the ETCC than to enzyme-substrate binding. However, in vitro phosphorylation and in vivo mutagenesis assays are required to verify this. Enzyme-substrate interactions may be transient and thus difficult to capture in proteomic analyses that are biased toward stable interactions, such as co-immunoprecipitation assays. The ability of the ETCC domain to form a more stable interaction with CK2 than otherwise allowed by the CK2 substrate regions may explain the BRD2/3/T-predominant CK2 associations described in the Lambert et al study[31]. However, as with PAF, the preference of CK2 for BRD2/3/T and the exact interacting fragments needs to be confirmed using purified proteins.

Potential roles for CK2 in regulating BET and PAF transcriptional pathways

CK2 is a ubiquitously expressed kinase complex with a vast array of targets, implicating it in numerous cellular functions and biological processes[216]. Given that it associates with many chromatin-bound proteins in yeast[167,217,218], it is not surprising that deletions of CK2 subunits result in global gene expression

changes[219]. In mammals, CK2 is preferentially bound to transcriptionally active nucleosomes[220], where it is thought to phosphorylate other nucleosome-bound proteins[221]. Candidate screens of potential CK2 phosphorylation targets have routinely yielded several chromatin-associated proteins, including members of the PAF complex and BRD2[215,222]. A screen for proteins bound to various chromatin fractions revealed that CK2 was associated with H3K9me3-marked chromatin[166], which may play a role in gene expression and chromatin organization[223]. In fact, the chromatin-bound fraction of CK2 is distributed maximally at gene promoters with extensions into gene bodies, particular at highly expressed genes, and with accumulation at enhancers, all of which suggest a role in transcription[168]. In yeast, CK2 was found to be associated with several transcription elongation factors including the PAF complex and IWS1[167]. In mammals, multiple CK2 target substrates were identified within PAF including 13 phosphorylation sites on the LEO1 component alone[222]. Interestingly, the physical association between CK2 and PAF is bridged by mutual interactions with the FACT complex[167,195], which is a histone chaperone that facilitates removal and replacement of the H2A/H2B dimer upon POL2 passage through the nucleosome[224]. Given that both FACT and PAF are required for optimal POL2 elongation, it is thought they may function in a coordinated and interdependent fashion[225]. Mechanistically, by recruiting CK2 to PAF, FACT promotes the phosphorylation of PAF, which is required for PAF-mediated H2B ubiquitylation[195]. Mutagenesis and structural studies are required to better understand how phosphorylation moderates PAF's function in transcription.

It will be particularly interesting to note how phosphorylation of PAF by CK2 affects PAF's ability to complex with elongating POL2.

The exact role of BRD2 (and potentially BRD3 and BRDT) in this picture remains elusive. Given that the biochemical studies identifying PAF, CK2, and FACT interactions were performed in yeast, it is possible that in mammals BRD2 replaces FACT as the mediator of PAF-CK2 contacts. Like FACT, BRD2 also possesses histone chaperone activity, though unlike FACT this activity is specific for hyperacetylated nucleosomes[40]. The shared functionality of BRD2 and FACT in these assays suggests common downstream mechanisms which may involve PAF and CK2. To clarify whether BRD2 recruits PAF and CK2 into close proximity, in vitro protein interactions studies should be performed using purified components. To assess in vivo, each factor should be targeted for depletion individually to reveal their roles on the others' ability to bind chromatin. For example, if BRD2 is playing a central role that governs CK2 phosphorylation of PAF and subsequent PAF-mediated POL2 elongation, depletion of BRD2 should result in reduced CK2 occupancy at the promoter and reduced levels of PAF and serine-2-phosphorylated POL2 throughout gene bodies. Development of phospho-PAF-specific antibodies would further help characterize distribution of the phosphorylated PAF in relation to total PAF on chromatin and the role of BRD2 and CK2 in this modification.

It is also possible that BRD2 interacts with PAF and CK2 separately, perhaps recruiting PAF to promoters (or facilitating its progression through nucleosomes) independently from any involvement in CK2-specific processes. For

example, CK2 has been implicated directly in transcriptional elongation via its ability to phosphorylate histone H2A at a conserved residue, which prevents H2B deubiquitylation by the SAGA complex[168]. It is therefore thought that CK2 regulates the opposing actions of PAF (H2B ubiquitylation) and SAGA (H2B deubiquitylation) to ultimately promote elongation via H2Bub. In this setting, BRD2 may separately recruit PAF and CK2 to active chromatin such that some BRD2 molecules help load PAF onto chromatin and into the elongating POL2 while others simultaneously give CK2 access to H2A substrates on acetylated nucleosomes. In this way, BRD2 may be functioning as a docking site for various elongation factors involved in different (or perhaps interdependent) aspects of transcriptional elongation. Additional structural characterization of the ETCC interactions will elucidate whether BRD2, PAF, and CK2 form a single complex or separate entities. Identification of a common binding motif recognized by the ETCC domain in subunits of both PAF and CK2 would support the notion of independent interactions.

In vivo, a large fraction of the chromatin bound CK2 is located at the promoter, and depending on the level of gene expression, CK2 signal is also detected throughout the gene body. PAF on the other hand has some signal at the promoter with stronger signal throughout the gene body and maximal signal at the termination site, largely paralleling that of elongating POL2 fraction marked by serine-2-phophorylation. BRD2 and the other BET proteins are mostly found at promoters. All three can be found distally at enhancers. The different distribution patterns of these three factors suggests that they do not form a single stable

complex on chromatin. Instead, the interactions we detect between the BRD2 ETCC domain and PAF and CK2 likely reflect momentary contacts that take place in vivo during transcription. For example, the bulk of BRD2 may be engaged with CK2 at the site of paused POL2. A smaller fraction of BRD2 may be present along the gene body and associate with PAF as it comes into contact with acetylated nucleosomes. Given that CK2 can phosphorylate BET proteins and various PAF subunits, we asked whether phosphorylation may be regulating the strength of the ETCC-PAF interaction. We found that in our pulldown assay the ETCC-PAF interaction was destabilized when phosphorylation was preserved through the use of phosphatase inhibitors. We therefore hypothesize that CK2 kinase activity regulates POL2 pause-release by releasing PAF from promoter-proximal BRD2 allowing it to incorporate into elongating POL2 (Figure 4.5.E).

However, much more work is needed to elaborate the details of this mechanism. First, as described in the prior section, the interface between the ETCC and the interacting PAF subunits needs to be identified. This structural characterization can then inform the search for the relevant residues that are phosphorylated by CK2. We expect that these amino acids will occur at an interface with the ETCC, but they could occur elsewhere and still exert intramolecular effects that modulate the ETCC interaction, akin to the phospho-switch mechanism that occurs at the motif-B proximal CK2 phosphorylation site in BET proteins[85]. It is also possible that the structurally relevant phosphorylated residues are not in fact CK2 substrates. CK2 is a strong candidate based on its physical association with PAF and the identification of CK2 targets therein, but

other promoter-proximal kinases may contribute. These candidates can be screened using kinase-specific small molecule inhibitors. The druggable kinases which are known to phosphorylate various transcription cofactors include CDK7[226–228], CDK9 (component of PTEFb, as described in previous sections)[229], CDK8/19[230], and CDK12/13[231]. Drugs targeting these kinases can be tested in parallel with CK2 inhibitors to identify the relevant factor that phosphorylates PAF and reduces its ability to bind the ETCC domain. In addition to confirming the correct kinase, we must also consider the opposing function of phosphatases. For example, the phosphatase PP2A has been shown to counter CK2-mediated phosphorylation of BRD4 such that the balance of CK2 and PP2A activity on BRD4 phosphorylation controls transcription and cell proliferation[86]. Chemical inhibition or activation of PP2A or other phosphatases can be tested to determine their role in controlling the ETCC-PAF interaction. Together, these studies will shed light on the molecular regulation of ETCC interactions.

Chemical or genetic perturbation of promoter-proximal kinases, CK2, and phosphatases may also be useful for studying the in vivo function of these factors in BET-mediated transcription. For example, it is well known that PTEFb inhibition (targeting CDK9 kinase catalytic activity) blocks POL2 pause-release resulting in reduced POL2 elongation and transcription[229]. Inhibition of BET proteins has largely the same effect as PTEFb blockade. Though this similarity is often attributed to BRD4's ability to recruit PTEFb[67,232], BET proteins can promote elongation independently from PTEFb[40,41,111,112]. Therefore, definitive molecular mechanisms implicating BET proteins in elongation remain to be identified. We

propose that BRD2/3 recruitment of PAF and CK2 may be one such mechanism. In support of this, the effect of CK2 inhibition on elongation largely reflects that of BET or PTEFb inhibition[168]. However, more detailed studies are required to examine how CK2 inhibition effects BET protein recruitment. For example, we can test whether CK2 activity is prerequisite for BET protein binding and more specifically whether BRD2/3 occupancy is more sensitive than BRD4 occupancy. It will also be interesting to note if CK2 activity is required for PAF recruitment to promoters or its release into gene bodies. In the latter case, CK2 inhibition may cause a pileup of PAF at the promoter. If so, we can then test if this promoter-proximal PAF is BRD2-dependent using BRD2 KO cells. This may reveal a role for BRD2 in recruiting PAF to chromatin before PAF is incorporated into elongating POL2 and relocated into gene bodies.

Unlike CK2 and PTEFb, the exact effect of PAF on POL2 pause-release and elongation remains controversial. However, it is possible that the contributions of BET proteins and CK2 could help reconcile these differences and reveal mechanistic insight into transcriptional regulation. For example, some studies show that PAF depletion mimics BET or CK2 perturbation by causing reduced elongation[130,197]. Other studies instead show an increase in elongation, suggesting that PAF actually promotes pausing by somehow tethering POL2 to the promoter perhaps in an enhancer-dependent fashion[198–200]. Interestingly, PAF and CK2 associate with FACT, which also has been found to exhibit dichotomous functions in transcription. FACT maintains POL2 in a promoter-proximal paused state on chromatin[233], even repressing the expression of genes where it sits at the

promoter[234,235], yet it also clearly functions as an elongation factor in transcription assays[224,236]. The distribution of FACT on chromatin resembles a blend of CK2 and PAF occupancy with a significant portion of FACT found at promoters as well as within gene bodies[233–235]. This enrichment over gene bodies is consistent with FACT's interactions with PAF as part of the elongating POL2 complex[164]. PAF and FACT may each have dual roles in pause-release and in elongation. Though BET proteins do not likely interact directly with FACT, they resemble FACT in transcription elongation assays and may associate together via PAF and CK2. Therefore, it is possible that BET proteins, alone or in conjunction with CK2, help to regulate or toggle between the various functions of PAF, potentially explaining some of the paradoxical effects when these factors are depleted.

The overlap between BET proteins, CK2, and PAF in various functional studies, protein interaction networks, and chromatin distribution profiles suggests a common transcriptional pathway that may be independent from or in conjunction with PTEFb-related elongation mechanisms. However, whereas PTEFb has a relatively straightforward role in pause-release, the BET-PAF-CK2 pathway may involve both pause-release and elongation. To test the role BRD2 in these pathways, the chromatin distributions of PAF and CK2 can be measured in BRD2 KO cells. To avoid the confounding effects of altered gene expression in the BRD2 KO cells, a rapid depletion system targeting BRD2 can be developed to better define the causal relationship between loss of BRD2 and the direct impact on the recruitment and distribution of the other factors. We have preliminary data that the auxin-inducible system can be used to promptly and reversibly degrade BRD2 (data not shown),

as has been done for BRD4[112]. Similar systems have been used to target PAF for depletion[200] and could be used in conjunction with CK2 kinase inhibition to elucidate the mechanics of this transcriptional pathway. Analysis of PAF and CK2 occupancy in BRD2 KO cells expressing BRD2 proteins with mutations in the ETCC regions will help determine the extent to which this domain is directly involved in PAF and CK2 recruitment.

In summary, we propose that BRD2 and BRD3 function independently from BRD4 through protein contacts with PAF and CK2. Though the mechanistic details remain unresolved, our data indicate that different elongation factors associate preferentially with individual BET proteins. These pathways are unique enough that BRD2 and BRD4 cannot genetically compensate for each other and thus operate within distinct functional niches on chromatin.

Implications for BET inhibitor studies and therapies

In this study we identified a transcriptional mechanism that may involve BRD2 and BRD3 but not BRD4. Classic BET inhibitors like the small molecules JQ1 or I-BET targeting the bromodomains typically lack specificity for individual BET proteins. Their impact on transcription is thus an aggregate effect. Based on our findings we expect that at least part of their effect can be attributed to impeding BRD2/3-dependent PAF and CK2 recruitment. Mechanistically, recruitment of PTEFb and subsequent POL2 pause-release is often considered the downstream effector of BET protein function. This activity is generally attributed to BRD4. Yet when BET proteins are perturbed pharmaceutically, POL2 elongation is impacted without disrupting PTEFb occupancy[111]. A BRD2/3-dependent elongation

mechanism involving CK2 and PAF may thus partly explain this PTEFb-independent elongation defect. It may also account for the different transcriptional signatures observed when comparing pan-BET and CDK9 (PTEFb) inhibitor treatment[112]. Furthermore, defective PAF and CK2 activity could account for the discrepancy between the effect of BRD4-specific depletion versus pan-BET bromodomain inhibition on gene expression[110]. In the future, careful comparison of genes affected by CK2 inhibition can help establish CK2 as a downstream mediator of BET proteins. Depletion studies can be performed to likewise identify common or distinct pathways involving PAF.

Though BD2-specific inhibitors target all of the BET proteins, they seem to affect BRD2 and BRD3 chromatin occupancy more so than BRD4 occupancy. Thus, BD2 inhibitors are functionally more specific for BRD2 and BRD3. In this setting, it was found that BD2 inhibition has greater specificity for inflammation-induced gene expression while leaving baseline gene expression alone[110]. This finding suggests that an intentionally targeted BRD2/3-based therapeutic may be better tolerated in patients. Toward this goal, our study proposes that the ETCC domain may offer a drug target that hits BRD2 and BRD3 while leaving BRD4 intact.

Concluding remarks

The results presented here provide a mechanistic rationale for distinct BET protein functions. Our study highlights the value of combining a powerful gene function assay with simple domain mapping experiments to drive discovery. Using this strategy, we demonstrate a framework for the development of BET-specific

inhibitors. And by identifying cofactors implicated in BET function, our study points toward new avenues of investigation to better understand how chromatin acetylation is linked to transcription.

BIBLIOGRAPHY

1. Werner MT, Wang H, Hamagami N, et al. Comparative structure-function analysis of bromodomain and extraterminal motif (BET) proteins in a gene-complementation system. *J Biol Chem*. 2020;295(7):1898-1914. doi:10.1074/jbc.RA119.010679
2. Olins AL, Olins DE. Spheroid Chromatin Units (v Bodies). *Science (80-)*. 1974;183(4122):330-332.
3. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol*. 2015;16(3):178-189. doi:10.1038/nrm3941
4. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A*. 1964;51(1938):786-794. doi:10.1073/pnas.51.5.786
5. Riggs MG, Whittaker RG, Neumann JR, Ingram VM. n-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells. *Nature*. 1977;268(5619):462-464. doi:10.1038/268462a0
6. Candido EPM, Reeves R, Davie JR. Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell*. 1978;14(1):105-113. doi:10.1016/0092-8674(78)90305-7
7. Vidali G, Boffa LC, Bradbury EM, Allfrey VG. Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequence. *Proc Natl Acad Sci U S A*. 1978;75(5):2239-2243. doi:10.1073/pnas.75.5.2239
8. Hebbes TR, Thorne AW, Crane-Robinson C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J*. 1988;7(5):1395-1402. doi:10.1002/j.1460-2075.1988.tb02956.x
9. Turner BM, Birley AJ, Lavender J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell*. 1992;69(2):375-384. doi:10.1016/0092-8674(92)90417-B
10. Brownell JE, Allis CD. An activity gel assay detects a single, catalytically active his tone acetyltransferase subunit in Tetrahymena macronuclei. *Proc Natl Acad Sci U S A*. 1995;92(14):6364-6368. doi:10.1073/pnas.92.14.6364
11. Taunton J, Hassig CA, Schreiber SL. A Mammalian Histone Deacetylase Related to the Yeast Transciptional Regulator Rpd3p. 1996;272(5260):408-411.
12. Davie JR, Chadee DN. Regulation and regulatory parameters of histone modifications. *J Cell Biochem*. 1998;72(S3031):203-213. doi:10.1002/(sici)1097-4644(1998)72:30/31+<203::aid-jcb25>3.3.co;2-w
13. Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*. 1999;399(6735):491-496. doi:10.1038/20974
14. Owen DJ, Ornaghi P, Yang J-C, et al. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J*. 2000;19(22):6141-6149. doi:10.1093/emboj/19.22.6141
15. Marmorstein R, Zhou MM. Writers and readers of histone acetylation: Structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol*. 2014;6(7). doi:10.1101/cshperspect.a018762
16. Clayton AL, Hazzalin CA, Mahadevan LC. Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. *Mol Cell*. 2006;23(3):289-296. doi:10.1016/j.molcel.2006.06.017
17. Verdin E, Ott M. 50 years of protein acetylation: From gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol*. 2015;16(4):258-264. doi:10.1038/nrm3931
18. Taniguchi Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. *Int J Mol Sci*. 2016;17. doi:10.3390/ijms17111849
19. Shang E, Salazar G, Crowley TE, et al. Identification of unique, differentiation stage-

specific patterns of expression of the bromodomain-containing genes Brd2, Brd3, Brd4, and Brdt in the mouse testis. *Gene Expr Patterns*. 2004;4(5):513-519. doi:10.1016/j.modgep.2004.03.002
20. Belkina AC, Denis G V. BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer*. 2012;12(7):465-477. doi:10.1038/nrc3256
21. Shang E, Wang X, Wen D, Greenberg DA, Wolgemuth DJ. Double bromodomain-containing gene Brd2 is essential for embryonic development in Mouse. *Dev Dyn*. 2009;238(4):908-917. doi:10.1002/dvdy.21911
22. Gyuris A, Donovan DJ, Seymour KA, et al. The chromatin-targeting protein Brd2 is required for neural tube closure and embryogenesis. *Biochim Biophys Acta*. 2009;1789(5):413-421. doi:10.1016/j.bbagrm.2009.03.005
23. Houzelstein D, Bullock SL, Lynch DE, Grigorieva EF, Wilson VA, Beddington RSP. Growth and Early Postimplantation Defects in Mice Deficient for the Bromodomain-Containing Protein Brd4. *Mol Cell Biol*. 2002;22(11):3794-3802. doi:10.1128/mcb.22.11.3794-3802.2002
24. Shang E, Nickerson HD, Wen D, Wang X, Wolgemuth DJ. The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development*. 2007;134(19):3507-3515. doi:10.1242/dev.004481
25. Denis G V., Green MR. A novel, mitogen-activated nuclear kinase is related to Drosophila developmental regulator. *Genes Dev*. 1996;10:261-271.
26. Florence B, McGinnis W. A genetic screen of the Drosophila X chromosome for mutations that modify Deformed function. *Genetics*. 1998;150(4):1497-1511.
27. Florence BL, Faller D V. Drosophila female sterile (1) homeotic is a multifunctional transcriptional regulator that is modulated by Ras signaling. *Dev Dyn*. 2008;237(3):554-564. doi:10.1002/dvdy.21432
28. Jiang YW, Veschambre P, Erdjument-Bromage H, et al. Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc Natl Acad Sci U S A*. 1998;95(15):8538-8543. doi:10.1073/pnas.95.15.8538
29. Denis G V., McComb ME, Faller D V., Sinha A, Romesser PB, Costello CE. Identification of transcription complexes that contain the double bromodomain protein Brd2 and chromatin remodeling machines. *J Proteome Res*. 2006;5(3):502-511. doi:10.1021/pr050430u
30. Dawson MA, Prinjha RK, Dittmann A, et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*. 2011;478(7370):529-533. doi:10.1038/nature10509
31. Lambert JP, Picaud S, Fujisawa T, et al. Interactome Rewiring Following Pharmacological Targeting of BET Bromodomains. *Mol Cell*. 2019;73(3):1-18. doi:10.1016/j.molcel.2018.11.006
32. Kanno T, Kanno Y, Siegel RM, Jang MK, Lenardo MJ, Ozato K. Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells. *Mol Cell*. 2004;13:33-43. doi:10.1038/srep38681
33. Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc Natl Acad Sci*. 2003;100(15):8758-8763. doi:10.1073/pnas.1433065100
34. Jung M, Philpott M, Müller S, et al. Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J Biol Chem*. 2014;289(13):9304-9319. doi:10.1074/jbc.M113.523019
35. Garcia-Gutierrez P, Mundi M, Garcia-Dominguez M. Association of bromodomain BET proteins with chromatin requires dimerization through the conserved motif B. *J Cell Sci*. 2012;125(15):3671-3680. doi:10.1242/jcs.105841
36. Hnilicova J, Hozeifi S, Stejskalova E, et al. The C-terminal domain of Brd2 is important for chromatin interaction and regulation of transcription and alternative splicing. *Mol Biol Cell*.

2013;24(22):3557-3568. doi:10.1091/mbc.e13-06-0303
37. Matangkasombut O, Buratowski RM, Swilling NW, Buratowski S. Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes Dev*. 2000;14(8):951-962.
38. Lygerou Z, Conesa C, Lesage P, et al. The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs. *Nucleic Acids Res*. 1994;22(24):5332-5340. doi:10.1093/nar/22.24.5332
39. Chang Y-L, King B, Lin S-C, Kennison JA, Huang D-H. A Double-Bromodomain Protein, FSH-S, Activates the Homeotic Gene Ultrabithorax through a Critical Promoter-Proximal Region. *Mol Cell Biol*. 2007;27(15):5486-5498. doi:10.1128/mcb.00692-07
40. LeRoy G, Rickards B, Flint SJ. The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription. *Mol Cell*. 2008;30(1):51-60. doi:10.1016/j.molcel.2008.01.018
41. Kanno T, Kanno Y, Leroy G, et al. BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. *Nat Struct Mol Biol*. 2014;21(12):1047-1057. doi:10.1038/nsmb.2912
42. Kasahara M, Hayashi M, Tanaka K, et al. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc Natl Acad Sci U S A*. 1996;93(17):9096-9101. doi:10.1073/pnas.93.17.9096
43. Thorpe KL, Gorman P, Thomas C, Sheer D, Trowsdale J, Beck S. Chromosomal localization, gene structure and transcription pattern of the ORFX gene, a homologue of the MHC-linked RING3 gene. *Gene*. 1997;200(1-2):177-183. doi:10.1016/S0378-1119(97)00415-0
44. Filippakopoulos P, Picaud S, Mangos M, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*. 2012;149(1):214-231. doi:10.1016/j.cell.2012.02.013
45. Smith SG, Zhou MM. The Bromodomain: A New Target in Emerging Epigenetic Medicine. *ACS Chem Biol*. 2016;11(3):598-608. doi:10.1021/acschembio.5b00831
46. Nakamura Y, Umehara T, Nakano K, et al. Crystal structure of the human BRD2 bromodomain: Insights into dimerization and recognition of acetylated histone H4. *J Biol Chem*. 2007;282(6):4193-4201. doi:10.1074/jbc.M605971200
47. Vollmuth F, Blankenfeldt W, Geyer M. Structures of the dual bromodomains of the P-TEFb-activating protein Brd4 at atomic resolution. *J Biol Chem*. 2009;284(52):36547-36556. doi:10.1074/jbc.M109.033712
48. Morinière J, Rousseaux S, Steuerwald U, et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*. 2009;461(7264):664-668. doi:10.1038/nature08397
49. Huang H, Zhang J, Shen W, et al. Solution structure of the second bromodomain of Brd2 and its specific interaction with acetylated histone tails. *BMC Struct Biol*. 2007;7:1-17. doi:10.1186/1472-6807-7-57
50. Gamsjaeger R, Webb SR, Lamonica JM, Billin A, Blobel GA, Mackay JP. Structural Basis and Specificity of Acetylated Transcription Factor GATA1 Recognition by BET Family Bromodomain Protein Brd3. *Mol Cell Biol*. 2011;31(13):2632-2640. doi:10.1128/mcb.05413-11
51. Umehara T, Nakamura Y, Jang MK, et al. Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J Biol Chem*. 2010;285(10):7610-7618. doi:10.1074/jbc.M109.062422
52. Umehara T, Nakamura Y, Wakamori M, Ozato K, Yokoyama S, Padmanabhan B. Structural implications for K5/K12-di-acetylated histone H4 recognition by the second bromodomain of BRD2. *FEBS Lett*. 2010;584(18):3901-3908. doi:10.1016/j.febslet.2010.08.013
53. Liu Y, Wang X, Zhang J, et al. Structural basis and binding properties of the second bromodomain of Brd4 with acetylated histone tails. *Biochemistry*. 2008;47(24):6403-6417. doi:10.1021/bi8001659

54. Miller TCR, Simon B, Rybin V, et al. A bromodomain-DNA interaction facilitates acetylation-dependent bivalent nucleosome recognition by the BET protein BRDT. *Nat Commun*. 2016;7. doi:10.1038/ncomms13855
55. LeRoy G, Chepelev I, DiMaggio PA, et al. Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. *Genome Biol*. 2012;13(8):R68. doi:10.1186/gb-2012-13-8-r68
56. Handoko L, Kaczkowski B, Hon C-C, et al. JQ1 affects BRD2-dependent and independent transcription regulation without disrupting H4-hyperacetylated chromatin states. *Epigenetics*. 2018;13(4):410-431. doi:10.1080/15592294.2018.1469891
57. Draker R, Ng MK, Sarcinella E, Ignatchenko V, Kislinger T, Cheung P. A Combination of H2A.Z and H4 Acetylation Recruits Brd2 to Chromatin during Transcriptional Activation. *PLoS Genet*. 2012;8(11):e1003047. doi:10.1371/journal.pgen.1003047
58. Vardabasso C, Gaspar-Maia A, Hasson D, et al. Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. *Mol Cell*. 2015;59(1):75-88. doi:10.1016/j.molcel.2015.05.009
59. Surface LE, Fields PA, Subramanian V, et al. H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain Poised Chromatin in ESCs. *Cell Rep*. 2016;14(5):1142-1155. doi:10.1016/j.celrep.2015.12.100
60. Shi J, Wang Y, Zeng L, et al. Disrupting the Interaction of BRD4 with Diacetylated Twist Suppresses Tumorigenesis in Basal-like Breast Cancer. *Cancer Cell*. 2014;25(2):210-225. doi:10.1016/j.ccr.2014.01.028
61. Roe JS, Mercan F, Rivera K, Pappin DJ, Vakoc CR. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol Cell*. 2015;58(6):1-12. doi:10.1016/j.molcel.2015.04.011
62. Huang B, Yang X-D, Zhou M-M, Ozato K, Chen L-F. Brd4 Coactivates Transcriptional Activation of NF- B via Specific Binding to Acetylated RelA. *Mol Cell Biol*. 2009;29(5):1375-1387. doi:10.1128/mcb.01365-08
63. Asangani IA, Dommeti VL, Wang X, et al. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature*. 2014;510(7504):278-282. doi:10.1038/nature13229
64. Rahman S, Sowa ME, Ottinger M, et al. The Brd4 Extraterminal Domain Confers Transcription Activation Independent of pTEFb by Recruiting Multiple Proteins, Including NSD3. *Mol Cell Biol*. 2011;31(13):2641-2652. doi:10.1128/mcb.01341-10
65. Shen C, Ipsaro JJ, Shi J, et al. NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. *Mol Cell*. 2015;60(6):847-859. doi:10.1016/j.molcel.2015.10.033
66. Zhang Q, Zeng L, Shen C, et al. Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4. *Structure*. 2016;24(7):1201-1208. doi:10.1016/j.str.2016.04.019
67. Liu W, Ma Q, Wong K, et al. Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell*. 2013;155(7):1581-1595. doi:10.1016/j.cell.2013.10.056
68. Crowe BL, Larue RC, Yuan C, Hess S, Kvaratskhelia M, Foster MP. Structure of the Brd4 ET domain bound to a C-terminal motif from γ-retroviral integrases reveals a conserved mechanism of interaction. *Proc Natl Acad Sci*. 2016;113(8):2086-2091. doi:10.1073/pnas.1516813113
69. Ottinger M, Christalla T, Nathan K, Brinkmann MM, Viejo-Borbolla A, Schulz TF. Kaposi's Sarcoma-Associated Herpesvirus LANA-1 Interacts with the Short Variant of BRD4 and Releases Cells from a BRD4- and BRD2/RING3-Induced G1 Cell Cycle Arrest. *J Virol*. 2006;80(21):10772-10786. doi:10.1128/jvi.00804-06
70. Conrad RJ, Fozouni P, Thomas S, et al. The Short Isoform of BRD4 Promotes HIV-1 Latency by Engaging Repressive SWI/SNF Chromatin-Remodeling Complexes. *Mol Cell*. 2017;67(6):1001-1012. doi:10.1016/j.molcel.2017.07.025
71. Wai DCC, Szyszka TN, Campbell AE, et al. The BRD3 ET domain recognizes a short

peptide motif through a mechanism that is conserved across chromatin remodelers and transcriptional regulators. *J Biol Chem*. 2018;293(19):7160-7175. doi:10.1074/jbc.RA117.000678
72. Lin Y-J, Umehara T, Inoue M, et al. Solution structure of the extraterminal domain of the bromodomain-containing protein BRD4. *Protein Sci*. 2008;17(12):2174-2179. doi:10.1110/ps.037580.108
73. Luna-Peláez N, March-Díaz R, Ceballos-Chávez M, et al. The Cornelia de Lange Syndrome-associated factor NIPBL interacts with BRD4 ET domain for transcription control of a common set of genes. *Cell Death Dis*. 2019;10(8):548. doi:10.1038/s41419-019-1792-x
74. Olley G, Ansari M, Bengani H, et al. BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome. *Nat Genet*. 2018;50(3):329-332. doi:10.1038/s41588-018-0042-y
75. Alesi V, Dentici ML, Loddo S, et al. Confirmation of BRD4 haploinsufficiency role in Cornelia de Lange–like phenotype and delineation of a 19p13.12p13.11 gene contiguous syndrome. *Ann Hum Genet*. 2019;83(2):100-109. doi:10.1111/ahg.12289
76. Bisgrove DA, Mahmoudi T, Henklein P, Verdin E. Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proc Natl Acad Sci*. 2007;104(34):13690-13695. doi:10.1073/pnas.0705053104
77. Li Y, Liu M, Chen LF, Chen R. P-TEFb: Finding its ways to release promoter-proximally paused RNA polymerase II. *Transcription*. 2018;9(2):88-94. doi:10.1080/21541264.2017.1281864
78. Yang Z, Yik JHN, Chen R, et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell*. 2005;19(4):535-545. doi:10.1016/j.molcel.2005.06.029
79. Devaiah BN, Case-Borden C, Gegonne A, et al. BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat Struct Mol Biol*. 2016;23(6):540-548. doi:10.1038/nsmb.3228
80. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*. 2015;33(6):661-667. doi:10.1038/nbt.3235
81. Alsarraj J, Walker RC, Webster JD, et al. Deletion of the proline-rich region of the murine metastasis susceptibility gene Brd4 promotes epithelial-to-mesenchymal transition- and stem cell-like conversion. *Cancer Res*. 2011;71(8):3121-3131. doi:10.1158/0008-5472.CAN-10-4417
82. Alsarraj J, Faraji F, Geiger TR, et al. BRD4 short isoform interacts with RRP1B, SIPA1 and components of the LINC complex at the inner face of the nuclear membrane. *PLoS One*. 2013;8(11):1-17. doi:10.1371/journal.pone.0080746
83. Luna-Peláez N, García-Domínguez M. Lyar-Mediated Recruitment of Brd2 to the Chromatin Attenuates Nanog Downregulation Following Induction of Differentiation. *J Mol Biol*. 2018;430(8):1084-1097. doi:10.1016/j.jmb.2018.02.023
84. Izumikawa K, Ishikawa H, Yoshikawa H, et al. LYAR potentiates rRNA synthesis by recruiting BRD2/4 and the MYST-type acetyltransferase KAT7 to rDNA. *Nucleic Acids Res*. 2019:1-16. doi:10.1093/nar/gkz747
85. Wu S-Y, Lee A-Y, Lai H-T, Zhang H, Chiang C-M. Phospho Switch Triggers Brd4 Chromatin Binding and Activator Recruitment for Gene-Specific Targeting. *Mol Cell*. 2013;49(5):843-857. doi:10.1016/j.molcel.2012.12.006
86. Shu S, Lin CY, He HH, et al. Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer. *Nature*. 2016;529(7586):413-417. doi:10.1038/nature16508
87. Shi J, Vakoc CR. The Mechanisms behind the Therapeutic Activity of BET Bromodomain Inhibition. *Mol Cell*. 2014;54(5):728-736. doi:10.1016/j.molcel.2014.05.016
88. Bolden JE, Tasdemir N, Dow LE, et al. Inducible in vivo silencing of Brd4 identifies potential toxicities of sustained BET protein inhibition. *Cell Rep*. 2014;8(6):1919-1929. doi:10.1016/j.celrep.2014.08.025

89. Belkina AC, Blanton WP, Nikolajczyk BS, Denis G V. The double bromodomain protein Brd2 promotes B cell expansion and mitogenesis. *J Leukoc Biol*. 2014;95(3):451-460. doi:10.1189/jlb.1112588
90. French C. NUT midline carcinoma. *Nat Rev Cancer*. 2014;14(3):149-150. doi:10.1038/nrc3659
91. Reynoird N, Schwartz BE, Delvecchio M, et al. Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *EMBO J*. 2010;29(17):2943-2952. doi:10.1038/emboj.2010.176
92. Alekseyenko AA, Walsh EM, Zee BM, et al. Ectopic protein interactions within BRD4-chromatin complexes drive oncogenic megadomain formation in NUT midline carcinoma. *Proc Natl Acad Sci U S A*. 2017;114(21):E4184-E4192. doi:10.1073/pnas.1702086114
93. Alekseyenko AA, Walsh EM, Wang X, et al. The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains. *Genes Dev*. 2015;29(14):1507-1523. doi:10.1101/gad.267583.115
94. Zuber J, Shi J, Wang E, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*. 2011;478(7370):524-528. doi:10.1038/nature10334
95. Filippakopoulos P, Qi J, Picaud S, et al. Selective inhibition of BET bromodomains. *Nature*. 2010;468(7327):1067-1073. doi:10.1038/nature09504
96. Nicodeme E, Jeffrey KL, Schaefer U, et al. Suppression of inflammation by a synthetic histone mimic. *Nature*. 2010;468(7327):1119-1123. doi:10.1038/nature09589
97. Hargreaves DC, Horng T, Medzhitov R. Control of Inducible Gene Expression by Signal-Dependent Transcriptional Elongation. *Cell*. 2009;138(1):129-145. doi:10.1016/j.cell.2009.05.047
98. Chapuy B, McKeown MR, Lin CY, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell*. 2013;24(6):777-790. doi:10.1016/j.ccr.2013.11.003
99. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153(2):320-334. doi:10.1016/j.cell.2013.03.036
100. Bai L, Zhou B, Yang CY, et al. Targeted degradation of BET proteins in triple-negative breast cancer. *Cancer Res*. 2017;77(9):2476-2487. doi:10.1158/0008-5472.CAN-16-2622
101. Hensel T, Giorgi C, Schmidt O, et al. Targeting the EWS-ETS transcriptional program by BET bromodomain inhibition in Ewing sarcoma. *Oncotarget*. 2015;7(2). doi:10.18632/oncotarget.6385
102. Chen Y, Xu L, Mayakonda A, et al. Bromodomain and extraterminal proteins foster the core transcriptional regulatory programs and confer vulnerability in liposarcoma. *Nat Commun*. 2019;10(1):1-14. doi:10.1038/s41467-019-09257-z
103. Xu L, Chen Y, Mayakonda A, et al. Targetable BET proteins- and E2F1-dependent transcriptional program maintains the malignancy of glioblastoma. *Proc Natl Acad Sci U S A*. 2018;115(22):E5086-E5095. doi:10.1073/pnas.1712363115
104. Deeney JT, Belkina AC, Shirihai OS, Corkey BE, Denis G V. BET Bromodomain proteins Brd2, Brd3 and Brd4 selectively regulate metabolic pathways in the pancreatic β-cell. *PLoS One*. 2016;11(3):1-16. doi:10.1371/journal.pone.0151329
105. Ding N, Hah N, Yu RT, et al. BRD4 is a novel therapeutic target for liver fibrosis. *Proc Natl Acad Sci U S A*. 2015;112(51):15713-15718. doi:10.1073/pnas.1522163112
106. Anand P, Brown JD, Lin CY, et al. BET bromodomains mediate transcriptional pause release in heart failure. *Cell*. 2013;154(3):569-582. doi:10.1016/j.cell.2013.07.013
107. Alqahtani A, Choucair K, Ashraf M, et al. Bromodomain and extra-terminal motif inhibitors: A review of preclinical and clinical advances in cancer therapy. *Futur Sci OA*. 2019;5(3). doi:10.4155/fsoa-2018-0115
108. Picaud S, Wells C, Felletar I, et al. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc Natl Acad Sci U S A*. 2013;110(49):19754-19759. doi:10.1073/pnas.1310658110
109. Faivre EJ, McDaniel KF, Albert DH, et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature*. 2020;578(7794):306-310. doi:10.1038/s41586-

020-1930-8
110. Gilan O, Rioja I, Knezevic K, et al. Selective targeting of BD1 and BD2 of the BET proteins in cancer and immuno-inflammation. *Science (80-)*. 2020. doi:10.1126/science.aaz8455
111. Winter GE, Mayer A, Buckley DL, et al. BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Mol Cell*. 2017;67(1):5-18. doi:10.1016/j.molcel.2017.06.004
112. Muhar M, Ebert A, Neumann T, et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science (80-)*. 2018;360(6390):800-805. doi:10.1126/science.aao2793
113. Stonestrom AJ, Hsu SC, Jahn KS, et al. Functions of BET proteins in erythroid gene expression. *Blood*. 2015;125(18):2825-2834. doi:10.1182/blood-2014-10-607309.The
114. Fong CY, Gilan O, Lam EYN, et al. BET inhibitor resistance emerges from leukaemia stem cells. *Nature*. 2015;525(7570):538-542. doi:10.1038/nature14888
115. Rathert P, Roth M, Neumann T, et al. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature*. 2015;525(7570):543-547. doi:10.1038/nature14898
116. Pawar A, Gollavilli PN, Wang S, Asangani IA. Resistance to BET Inhibitor Leads to Alternative Therapeutic Vulnerabilities in Castration-Resistant Prostate Cancer. *Cell Rep*. 2018;22(9):2236-2245. doi:10.1016/j.celrep.2018.02.011
117. Anders L, Guenther MG, Qi J, et al. Genome-wide localization of small molecules. *Nat Biotechnol*. 2014;32(1):92-96. doi:10.1038/nbt.2776
118. Khoueiry P, Ward Gahlawat A, Petretich M, et al. BRD4 bimodal binding at promoters and drug-induced displacement at Pol II pause sites associates with I-BET sensitivity. *Epigenetics Chromatin*. 2019;12(1):39. doi:10.1186/s13072-019-0286-5
119. Wang F, Liu H, Blanton WP, Belkina A, Lebrasseur NK, Denis G V. Brd2 disruption in mice causes severe obesity without Type 2 diabetes. *Biochem J*. 2009;425(1):71-83. doi:10.1042/bj20090928
120. Gollavilli PN, Pawar A, Wilder-Romans K, et al. Ews/ets-driven ewing sarcoma requires bet bromodomain proteins. *Cancer Res*. 2018;78(16):4760-4773. doi:10.1158/0008-5472.CAN-18-0484
121. Loganathan SN, Tang N, Fleming JT, et al. BET bromodomain inhibitors suppress EWS-FLI1-dependent transcription and the IGF1 autocrine mechanism in Ewing sarcoma. *Oncotarget*. 2016;7(28). doi:10.18632/oncotarget.9762
122. Cheung KL, Zhang F, Jaganathan A, et al. Distinct Roles of Brd2 and Brd4 in Potentiating the Transcriptional Program for Th17 Cell Differentiation. *Mol Cell*. 2017;65(6):1-13. doi:10.1016/j.molcel.2016.12.022
123. Belkina AC, Nikolajczyk BS, Denis G V. BET Protein Function Is Required for Inflammation: Brd2 Genetic Disruption and BET Inhibitor JQ1 Impair Mouse Macrophage Inflammatory Responses. *J Immunol*. 2013;190(7):3670-3678. doi:10.4049/jimmunol.1202838
124. Andrieu GP, Denis G V. BET Proteins Exhibit Transcriptional and Functional Opposition in the Epithelial-to-Mesenchymal Transition. *Mol Cancer Res*. 2018;16(4):580-586. doi:10.1158/1541-7786.mcr-17-0568
125. Roberts TC, Etxaniz U, Dall'agnese A, et al. BRD3 and BRD4 BET Bromodomain Proteins Differentially Regulate Skeletal Myogenesis. *Sci Rep*. 2017;7(1):1-16. doi:10.1038/s41598-017-06483-7
126. Fernandez-Alonso R, Davidson L, Hukelmann J, et al. Brd4-Brd2 isoform switching coordinates pluripotent exit and Smad2-dependent lineage specification. *EMBO Rep*. 2017;18(7):1108-1122. doi:10.15252/embr.201643534
127. Gursoy-Yuzugullu O, Carman C, Price BD. Spatially restricted loading of BRD2 at DNA double-strand breaks protects H4 acetylation domains and promotes DNA repair. *Sci Rep*. 2017;7(1):1-13. doi:10.1038/s41598-017-13036-5
128. Floyd SR, Pacold ME, Huang Q, et al. The bromodomain protein Brd4 insulates chromatin from DNA damage signalling. *Nature*. 2013;498(7453):246-250. doi:10.1038/nature12147

129. Fontanals-Cirera B, Hasson D, Vardabasso C, et al. Harnessing BET Inhibitor Sensitivity Reveals AMIGO2 as a Melanoma Survival Gene. *Mol Cell*. 2017;68(4):731-744.e9. doi:10.1016/j.molcel.2017.11.004
130. Yu M, Yang W, Ni T, et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science (80-)*. 2015;350(6266):1383-1386. doi:10.1126/science.aad2338
131. Hsu SC, Gilgenast TG, Bartman CR, et al. The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. *Mol Cell*. 2017;66(1):102-116. doi:10.1016/j.molcel.2017.02.027
132. Stonestrom AJ, Hsu SC, Werner MT, Blobel GA. Erythropoiesis provides a BRD's eye view of BET protein function. *Drug Discov Today Technol*. 2016;19:23-28. doi:10.1016/j.ddtec.2016.05.004
133. Berthon C, Raffoux E, Thomas X, et al. Bromodomain inhibitor OTX015 in patients with acute leukaemia: a dose-escalation, phase 1 study. *Lancet Haematol*. 2016;3(4):e186-e195. doi:10.1016/s2352-3026(15)00247-1
134. Amorim S, Stathis A, Gleeson M, et al. Bromodomain inhibitor OTX015 in patients with lymphoma or multiple myeloma: A dose-escalation, open-label, pharmacokinetic, phase 1 study. *Lancet Haematol*. 2016;3(4):e196-e204. doi:10.1016/S2352-3026(16)00021-1
135. Piha-Paul SA, Sachdev JC, Barve MA, et al. Journal of Clinical Oncology. *J Clin Oncol*. 2018;36(15_suppl):2510. doi:10.1200/JCO.2018.36.15_suppl.2510
136. Ferreira R, Ohneda K, Yamamoto M, Philipsen S. GATA1 Function , a Paradigm for Transcription Factors in Hematopoiesis. *Mol Cell Biol*. 2005;25(4):1215-1227. doi:10.1128/MCB.25.4.1215
137. Boyes J, Byfield P, Nakatani Y, Ogryzko V. Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature*. 1998;396:594-598. doi:10.1038/255243a0
138. Hung H-L, Lau J, Kim AY, Weiss MJ, Blobel GA. CREB-Binding Protein Acetylates Hematopoietic Transcription Factor GATA-1 at Functionally Important Sites. *Mol Cell Biol*. 1999;19(5):3496-3505. doi:10.1128/mcb.19.5.3496
139. Lamonica JM, Deng W, Kadauke S, et al. Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proc Natl Acad Sci*. 2011;108(22):E159-E168. doi:10.1073/pnas.1102140108
140. Lamonica JM, Vakoc CR, Blobel GA. Acetylation of GATA-1 is required for chromatin occupancy. *Blood*. 2006;108(12):3736-3738. doi:10.1182/blood-2006-07-032847
141. Weiss MJ, Yu C, Orkin SH. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol*. 1997;17(3):1642-1651. doi:10.1128/mcb.17.3.1642
142. Gregory T, Yu C, Ma A, Orkin SH, Blobel GA, Weiss MJ. GATA-1 and erythropoietin cooperate to promote erythroid cell survival by regulating bcl-xL expression. *Blood*. 1999;94(1):87-96. http://www.ncbi.nlm.nih.gov/pubmed/10381501.
143. Delorenzi M, Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*. 2002;18(4):617-625. doi:10.1093/bioinformatics/18.4.617
144. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191. doi:10.1093/bioinformatics/btp033
145. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127. doi:10.1093/biostatistics/kxj037
146. Jang MK, Mochizuki K, Zhou M, Jeong H, Brady JN, Ozato K. The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription. *Mol Cell*. 2005;19:523-534. doi:10.1016/j.molcel.2005.06.027
147. Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*. 2004;104(10):3136-3147. doi:10.1182/blood-2004-04-1603

148. Paralkar VR, Mishra T, Luan J, et al. Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. 2014;123(12):1927-1938. doi:10.1182/blood-2013-12-544494.The
149. Chung J, Chen C, Paw BH. Heme metabolism and erythropoiesis. *Curr Opin Hematol*. 2012;19(3):156-162. doi:10.1097/MOH.0b013e328351c48b
150. Fan J, Rone MB, Papadopoulos V. Translocator protein 2 is involved in cholesterol redistribution during erythropoiesis. *J Biol Chem*. 2009;284(44):30484-30497. doi:10.1074/jbc.M109.029876
151. Knight ZA, Schmidt SF, Birsoy K, Tan K, Friedman JM. A critical role for mTORC1 in erythropoiesis and anemia. *Elife*. 2014;3:1-17. doi:10.7554/elife.01913
152. Devaiah BN, Lewis BA, Cherman N, et al. BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain. *Proc Natl Acad Sci*. 2012;109(18):6927-6932. doi:10.1073/pnas.1120422109
153. Guo N, Faller D V, Denis G V. Activation-induced nuclear translocation of RING3. *J Cell Sci*. 2000;113:3085-3091. http://www.ncbi.nlm.nih.gov/pubmed/10934046%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3936601.
154. Fukazawa H, Masumi A. The Conserved 12-Amino Acid Stretch in the Inter-Bromodomain Region of BET Family Proteins Functions as a Nuclear Localization Signal. *Biol Pharm Bull*. 2012;35(11):2064-2068. doi:10.1248/bpb.b12-00527
155. Jain D, Mishra T, Giardine BM, et al. Dynamics of GATA1 binding and expression response in a GATA1-induced erythroid differentiation system. *Genomics Data*. 2015;4:1-7. doi:10.1016/j.gdata.2015.01.008
156. Zhang P, Wang D, Zhao Y, et al. Intrinsic BET inhibitor resistance in SPOP-mutated prostate cancer is mediated by BET protein stabilization and AKT-mTORC1 activation. *Nat Med*. 2017;23(9):1055-1062. doi:10.1038/nm.4379
157. Janouskova H, El Tekle G, Bellini E, et al. Opposing effects of cancer-Type-specific SPOP mutants on BET protein degradation and sensitivity to BET inhibitors. *Nat Med*. 2017;23(9):1046-1054. doi:10.1038/nm.4372
158. Dai X, Gan W, Li X, et al. Prostate cancer-Associated SPOP mutations confer resistance to BET inhibitors through stabilization of BRD4. *Nat Med*. 2017;23(9):1063-1071. doi:10.1038/nm.4378
159. Truebestein L, Leonard TA. Coiled-coils: The long and short of it. *BioEssays*. 2016;38(9):903-916. doi:10.1002/bies.201600062
160. Di Micco R, Fontanals-Cirera B, Low V, et al. Control of Embryonic Stem Cell Identity by BRD4-Dependent Transcriptional Elongation of Super-Enhancer-Associated Pluripotency Genes. *Cell Rep*. 2014;9(1):234-247. doi:10.1016/j.celrep.2014.08.055
161. Schröder S, Cho S, Zeng L, et al. Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. *J Biol Chem*. 2012;287(2):1090-1099. doi:10.1074/jbc.M111.282855
162. C.C.-S. H, C.R. B, P. H, et al. A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. *Genes Dev*. 2016;30(12):1423-1439. doi:http://dx.doi.org/10.1101/gad.280859
163. Doré LC, Crispino JD. Transcription factor networks in erythroid cell and megakaryocyte development. *Blood*. 2011;118(2):231-239. doi:10.1182/blood-2011-04-285981
164. Vos SM, Farnung L, Boehning M, et al. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*. 2018;560(7720):607-612. doi:10.1038/s41586-018-0440-4
165. Kim J, Guermah M, Roeder RG. The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. *Cell*. 2010;140(4):491-503. doi:10.1016/j.cell.2009.12.050
166. Ji X, Dadon DB, Abraham BJ, et al. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc Natl Acad Sci U S A*. 2015;112(12):3841-3846. doi:10.1073/pnas.1502971112
167. Krogan NJ, Kim M, Ahn SH, et al. RNA Polymerase II Elongation Factors of

Saccharomyces cerevisiae: a Targeted Proteomics Approach. *Mol Cell Biol*. 2002;22(20):6979-6992. doi:10.1128/mcb.22.20.6979-6992.2002
168. Basnet H, Su XB, Tan Y, et al. Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. *Nature*. 2014;516(7530):267-271. doi:10.1038/nature13736
169. Yang Y, Li W, Hoque M, et al. PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. *PLoS Genet*. 2016;12(1):1-28. doi:10.1371/journal.pgen.1005794
170. Van Oss SB, Shirra MK, Bataille AR, et al. The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6. *Mol Cell*. 2016;64(4):815-825. doi:10.1016/j.molcel.2016.10.008
171. Federation AJ, Nandakumar V, Searle BC, et al. Highly Parallel Quantification and Compartment Localization of Transcription Factors and Nuclear Proteins. *Cell Rep*. 2020;30(8):2463-2471.e5. doi:10.1016/j.celrep.2020.01.096
172. Bhagwat AS, Roe JS, Mok BYL, Hohmann AF, Shi J, Vakoc CR. BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. *Cell Rep*. 2016;15(3):519-530. doi:10.1016/j.celrep.2016.03.054
173. Malovannaya A, Li Y, Bulynko Y, et al. Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc Natl Acad Sci U S A*. 2010;107(6):2431-2436. doi:10.1073/pnas.0912599106
174. Malovannaya A, Lanz RB, Jung SY, et al. Analysis of the human endogenous coregulator complexome. *Cell*. 2011;145(5):787-799. doi:10.1016/j.cell.2011.05.006
175. Ghosh K, Tang M, Kumari N, et al. Positive Regulation of Transcription by Human ZMYND8 through Its Association with P-TEFb Complex. *Cell Rep*. 2018;24(8):2141-2154.e6. doi:10.1016/j.celrep.2018.07.064
176. Nishibuchi G, Shibata Y, Hayakawa T, et al. Physical and functional interactions between the histone H3K4 demethylase KDM5A and the nucleosome remodeling and deacetylase (NuRD) complex. *J Biol Chem*. 2014;289(42):28956-28970. doi:10.1074/jbc.M114.573725
177. Shiota H, Elya JE, Alekseyenko AA, et al. “Z4” complex member fusions in nut carcinoma: Implications for a novel oncogenic mechanism. *Mol Cancer Res*. 2018;16(12):1826-1833. doi:10.1158/1541-7786.MCR-18-0474
178. Gong F, Chiu LY, Cox B, et al. Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. *Genes Dev*. 2015;29(2):197-211. doi:10.1101/gad.252189.114
179. Gong F, Clouaire T, Aguirrebengoa M, Legube G, Miller KM. Histone demethylase KDM5A regulates the ZMYND8-NuRD chromatin remodeler to promote DNA repair. *J Cell Biol*. 2017;216(7):1959-1974. doi:10.1083/jcb.201611135
180. Sun Z, Feng D, Everett LJ, Bugge A, Lazar MA. Circadian epigenomic remodeling and hepatic lipogenesis: Lessons from HDAC3. *Cold Spring Harb Symp Quant Biol*. 2011;76:49-55. doi:10.1101/sqb.2011.76.011494
181. Han X, Yu D, Gu R, et al. Roles of the BRD4 short isoform in phase separation and active gene transcription. *Nat Struct Mol Biol*. 2020:1-9. doi:10.1038/s41594-020-0394-8
182. Crawford NPS, Alsarraj J, Lukes L, et al. Bromodomain 4 activation predicts breast cancer survival. *Proc Natl Acad Sci U S A*. 2008;105(17):6380-6385. doi:10.1073/pnas.0710331105
183. Pivot-pajot C, Caron C, Govin J, Vion A, Rousseaux S, Khochbin S. Acetylation-Dependent Chromatin Reorganization by BRDT , a Testis-Specific Bromodomain-Containing Protein. *Mol Cell Biol*. 2003;23(15):5354-5365. doi:10.1128/MCB.23.15.5354
184. Berkovits BD, Wolgemuth DJ. The first bromodomain of the testis-specific double bromodomain protein Brdt is required for chromocenter organization that is modulated by genetic background. *Dev Biol*. 2011;360(2):358-368. doi:10.1016/j.ydbio.2011.10.005
185. Dhar S, Thota A, Rao MRS. Insights into role of bromodomain, testis-specific (Brdt) in acetylated histone H4-dependent chromatin remodeling in mammalian spermiogenesis. *J Biol Chem*. 2012;287(9):6387-6405. doi:10.1074/jbc.M111.288167
186. Gaucher J, Boussouar F, Montellier E, et al. Bromodomain-dependent stage-specific male

genome programming by Brdt. *EMBO J*. 2012;31(19):3809-3820. doi:10.1038/emboj.2012.233
187. Ladurner AG, Inouye C, Jain R, Tjian R. Bromodomains Mediate an Acetyl-Histone Encoded Antisilencing Function at Heterochromatin Boundaries. *Mol Cell*. 2003;11:365-376. http://www.molecule.org/cgi/.
188. Wang R, Li Q, Helfer CM, Jiao J, You J. Bromodomain protein Brd4 associated with acetylated chromatin is important for maintenance of higher-order chromatin structure. *J Biol Chem*. 2012;287(14):10738-10752. doi:10.1074/jbc.M111.323493
189. Bryant JM, Donahue G, Wang X, et al. Characterization of BRD4 during Mammalian Postmeiotic Sperm Development. *Mol Cell Biol*. 2015;35(8):1433-1448. doi:10.1128/mcb.01328-14
190. Matzuk MM, McKeown MR, Filippakopoulos P, et al. Small-molecule inhibition of BRDT for male contraception. *Cell*. 2012;150(4):673-684. doi:10.1016/j.cell.2012.06.045
191. Kimura S, Loppin B. Two bromodomain proteins functionally interact to recapitulate an essential BRDT-like function in Drosophila spermatocytes. *Open Biol*. 2015;5(2):0-9. doi:10.1098/rsob.140145
192. Mackay JP, Sunde M, Lowry JA, Crossley M, Matthews JM. Protein interactions: is seeing believing? *Trends Biochem Sci*. 2007;32(12):530-531. doi:10.1016/j.tibs.2007.09.006
193. Xu Y, Bernecky C, Lee CT, et al. Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nat Commun*. 2017;8. doi:10.1038/ncomms15741
194. Torrado M, Low JKK, Silva APG, et al. Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex. *FEBS J*. 2017;284(24):4216-4232. doi:10.1111/febs.14301
195. Bedard LG, Dronamraju R, Kerschner JL, et al. Quantitative analysis of dynamic protein interactions during transcription reveals a role for casein kinase II in Polymerase-associated Factor (PAF) complex phosphorylation and regulation of histone H2B monoubiquitylation. *J Biol Chem*. 2016;291(26):13410-13420. doi:10.1074/jbc.M116.727735
196. Vos SM, Farnung L, Urlaub H, Cramer P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*. 2018;560(7720):601-606. doi:10.1038/s41586-018-0442-2
197. Hou L, Wang Y, Liu Y, et al. Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc Natl Acad Sci U S A*. 2019;116(29):14583-14592. doi:10.1073/pnas.1904324116
198. Bai X, Kim J, Yang Z, et al. TIF1γ Controls Erythroid Cell Fate by Regulating Transcription Elongation. *Cell*. 2010;142(1):133-143. doi:10.1016/j.cell.2010.05.028
199. Chen FX, Woodfin AR, Gardini A, et al. PAF1 , a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II Article PAF1 , a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell*. 2015;162(5):1003-1015. doi:10.1016/j.cell.2015.07.042
200. Chen FX, Xie P, Collings CK, et al. PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science (80-)*. 2017;357(6357):1294-1298. doi:10.1126/science.aan3269
201. Fischl H, Howe FS, Furger A, Mellor J. Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate. *Mol Cell*. 2017;65(4):685-698.e8. doi:10.1016/j.molcel.2017.01.006
202. Van Oss SB, Cucinotta CE, Arndt KM. Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. *Trends Biochem Sci*. 2017;42(10):788-798. doi:10.1016/j.tibs.2017.08.003
203. Ellison MA, Lederer AR, Warner MH, et al. The Paf1 Complex Broadly Impacts the Transcriptome of Saccharomyces cerevisiae. *Genetics*. 2019;212(3):711-728. doi:10.1534/genetics.119.302262
204. Lu X, Zhu X, Li Y, et al. Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. *Nucleic Acids Res*. 2016;44(14):6853-6867. doi:10.1093/nar/gkw571

205. He N, Chan CK, Sobhian B, et al. Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. *Proc Natl Acad Sci U S A*. 2011;108(36). doi:10.1073/pnas.1107107108
206. Wan L, Chong S, Xuan F, et al. Impaired cell fate through gain-of-function mutations in a chromatin reader. *Nature*. 2020;577(7788):121-126. doi:10.1038/s41586-019-1842-7
207. Hanks S, Perdeaux ER, Seal S, et al. Germline mutations in the PAF1 complex gene CTR9 predispose to Wilms tumour. *Nat Commun*. 2014;5:1-7. doi:10.1038/ncomms5398
208. Jurynec MJ, Bai X, Bisgrove BW, et al. The Paf1 complex and P-TEFb have reciprocal and antagonist roles in maintaining multipotent neural crest progenitors. *Dev*. 2019;146(24):1-12. doi:10.1242/dev.180133
209. Ding L, Paszkowski-Rogacz M, Nitzsche A, et al. A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. *Cell Stem Cell*. 2009;4(5):403-415. doi:10.1016/j.stem.2009.03.009
210. Tastemel M, Gogate AA, Malladi VS, et al. Transcription pausing regulates mouse embryonic stem cell differentiation. *Stem Cell Res*. 2017;25:250-255. doi:10.1016/j.scr.2017.11.012
211. Kim S, Kim JD, Chung AY, et al. Antagonistic regulation of PAF1C and p-TEFb is required for oligodendrocyte differentiation. *J Neurosci*. 2012;32(24):8201-8207. doi:10.1523/JNEUROSCI.5344-11.2012
212. Wu L, Li L, Zhou B, Qin Z, Dou Y. H2B Ubiquitylation Promotes RNA Pol II Processivity via PAF1 and pTEFb. *Mol Cell*. 2014;54(6):920-931. doi:10.1016/j.molcel.2014.04.013
213. Wu SY, Nin DS, Lee AY, Simanski S, Kodadek T, Chiang CM. BRD4 Phosphorylation Regulates HPV E2-Mediated Viral Transcription, Origin Replication, and Cellular MMP-9 Expression. *Cell Rep*. 2016;16(6):1733-1748. doi:10.1016/j.celrep.2016.07.001
214. Sawa C, Nedea E, Krogan N, et al. Bromodomain Factor 1 (Bdf1) Is Phosphorylated by Protein Kinase CK2. *Mol Cell Biol*. 2004;24(11):4734-4742. doi:10.1128/mcb.24.11.4734-4742.2004
215. Varjosalo M, Sacco R, Stukalov A, et al. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nat Methods*. 2013;10(4):307-314. doi:10.1038/nmeth.2400
216. Litchfield DW. Protein kinase CK2: Structure, regulation and role in cellular decisions of life and death. *Biochem J*. 2003;369(1):1-15. doi:10.1042/BJ20021469
217. Gavin AC, Bösche M, Krause R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*. 2002;415(6868):141-147. doi:10.1038/415141a
218. Ho Y, Ho Y, Gruhler A, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*. 2002;415(6868):180-183. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11805837%5Cnpapers3://publication/uuid/BB1850BC-BC4C-4AF7-9352-F1FF171C4B68.
219. Barz T, Ackermann K, Dubois G, Eils R, Pyerin W. Genome-wide expression screens indicate a global role for protein kinase CK2 in chromatin remodeling. *J Cell Sci*. 2003;116(8):1563-1577. doi:10.1242/jcs.00352
220. Guo C, Davis AT, Ahmed K. Dynamics of protein kinase CK2 association with nucleosomes in relation to transcriptional activity. *J Biol Chem*. 1998;273(22):13675-13680. doi:10.1074/jbc.273.22.13675
221. Guo C, Davis AT, Yu S, Tawfic S, Ahmed K. Role of protein kinase CK2 in phosphorylation of nucleosomal proteins in relation to transcriptional activity. *Mol Cell Biochem*. 1999;191(1-2):135-142. doi:10.1023/A:1006881405383
222. Bian Y, Ye M, Wang C, et al. Global screening of CK2 kinase substrates by an integrated phosphoproteomics workflow. *Sci Rep*. 2013;3:1-7. doi:10.1038/srep03460
223. Rosenfeld JA, Xuan Z, DeSalle R. Investigating repetitively matching short sequencing reads: The enigmatic nature of H3K9me3. *Epigenetics*. 2009;4(7):476-486. doi:10.4161/epi.4.7.9809

224. Hsieh FK, Kulaeva OI, Patel SS, et al. Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc Natl Acad Sci U S A*. 2013;110(19):7654-7659. doi:10.1073/pnas.1222198110
225. Pavri R, Zhu B, Li G, et al. Histone H2B Monoubiquitination Functions Cooperatively with FACT to Regulate Elongation by RNA Polymerase II. *Cell*. 2006;125(4):703-717. doi:10.1016/j.cell.2006.04.029
226. Glover-Cutter K, Larochelle S, Erickson B, et al. TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Mol Cell Biol*. 2009;29(20):5455-5464. doi:10.1128/mcb.00637-09
227. Ur Rasool R, Natesan R, Deng Q, et al. CDK7 inhibition suppresses castration-resistant prostate cancer through MED1 inactivation. *Cancer Discov*. 2019;9(11):1538-1555. doi:10.1158/2159-8290.CD-19-0189
228. Joo YJ, Ficarro SB, Chun Y, Marto JA, Buratowski S. In vitro analysis of RNA polymerase II elongation complex dynamics. *Genes Dev*. 2019;33(9-10):578-589. doi:10.1101/gad.324202.119
229. Bacon CW, D'Orso I. CDK9: a signaling hub for transcriptional control. *Transcription*. 2019;10(2):57-75. doi:10.1080/21541264.2018.1523668
230. Fant CB, Taatjes DJ. Regulatory functions of the Mediator kinases CDK8 and CDK19. *Transcription*. 2019;10(2):76-90. doi:10.1080/21541264.2018.1556915
231. Greenleaf AL. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. *Transcription*. 2019;10(2):91-110. doi:10.1080/21541264.2018.1535211
232. Patel MC, Debrosse M, Smith M, et al. BRD4 Coordinates Recruitment of Pause Release Factor P-TEFb and the Pausing Complex NELF/DSIF To Regulate Transcription Elongation of Interferon-Stimulated Genes. *Mol Cell Biol*. 2013;33(12):2497-2507. doi:10.1128/mcb.01180-12
233. Tettey TT, Gao X, Shao W, et al. A Role for FACT in RNA Polymerase II Promoter-Proximal Pausing. *Cell Rep*. 2019;27(13):3770-3779.e7. doi:10.1016/j.celrep.2019.05.099
234. Mylonas C, Tessarz P. Transcriptional repression by FACT is linked to regulation of chromatin accessibility at the promoter of ES cells. *Life Sci Alliance*. 2018;1(3):1-14. doi:10.26508/lsa.201800085
235. Kolundzic E, Ofenbauer A, Bulut SI, et al. FACT Sets a Barrier for Cell Fate Reprogramming in Caenorhabditis elegans and Human Cells. *Dev Cell*. 2018;46(5):611-626.e12. doi:10.1016/j.devcel.2018.07.006
236. Orphanides G, LeRoy G, Chang CH, Luse DS, Reinberg D. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell*. 1998;92(1):105-116. doi:10.1016/S0092-8674(00)80903-4